CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Identification of Phytochemicals from *Pennisetum glaucum* as Potential Inhibitors of EGFR in Non-Small Cell Lung Cancer

by

Amna Nawaz

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

I dedicate this thesis to my beloved father, whose unwavering support and prayers have been my strength, and to Dr. Erum Dilshad, whose guidance and inspiration have been invaluable throughout this journey.

CERTIFICATE OF APPROVAL

Identification of Phytochemicals from *Pennisetum glaucum* as Potential Inhibitors of EGFR in Non-Small Cell Lung Cancer

by

Amna Nawaz

(MBS233012)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Sara Mumtaz	NUMS, Rawalpindi
(b)	Internal Examiner	Dr. Muhammad Asad Anwar	CUST, Islamabad
(c)	Supervisor	Dr. Erum Dilshad	CUST, Islamabad

Dr. Erum Dilshad
Thesis Supervisor
April, 2025

Dr. Syeda Marriam Bakhtiar
Head
Dept. of BI and BS
April, 2025

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
April, 2025

Author's Declaration

I, **Amna Nawaz** hereby state that my MS thesis titled "**Identification of Phytochemicals from *Pennisetum glaucum* as Potential Inhibitors of EGFR in Non-Small Cell Lung Cancer**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

Amna Nawaz

Registration No: MBS233012

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "Identification of Phytochemicals from *Pennisetum glaucum* as Potential Inhibitors of EGFR in Non-Small Cell Lung Cancer" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Amna Nawaz)

Registration No: MBS233012

Acknowledgement

All praise and gratitude to Allah Almighty for His countless blessings, guidance, and strength throughout this journey. I am deeply thankful to my family, especially my father, for their unwavering love, prayers, and encouragement, which have been my greatest source of motivation. I extend my heartfelt appreciation to my supervisor, Dr. Erum Dilshad, for her invaluable guidance, support, and inspiration, which played a crucial role in the completion of this work. Special thanks to Sumaira for her constant encouragement, assistance, and belief in me during challenging times. This achievement would not have been possible without all of your contributions.

(Amna Nawaz)

Abstract

Non-Small Cell Lung Cancer (NSCLC) is the most common type of lung cancer, accounting for approximately 80 - 85% of all cases. NSCLC includes subtypes such as adenocarcinoma, squamous cell carcinoma, and large cell carcinoma, each with distinct histological characteristics. Despite advancements in treatments like targeted therapies and immunotherapy, NSCLC remains a leading cause of cancer-related mortality worldwide.

The aim of this study was to identify phytocompounds (from *Pennisetum glaucum*) as potential inhibitors of EGFR in NSCLC. A total of fifteen compounds were selected for docking against the EGFR protein. The compounds underwent virtual screening based on Lipinski's Rule of Five and ADMET analysis using the PkCSM tool. Docking simulations were then performed using CB-Dock, and the docking results were visualized with LIGPLOT Plus.

Among the tested compounds, acacetin emerged as the lead candidate due to its compliance with Lipinski's Rule of Five, favorable ADMET and physicochemical properties, and superior hydrophobic interactions compared to gefitinib. Given the success of the virtual drug design, further in vitro and in vivo studies are recommended to validate acacetin's potential as an EGFR inhibitor in NSCLC.

Keywords: NSCLC, Phytocompunds, Drug development, Acacetin, *Pennisetum glaucum.*

Contents

List of Figures

List of Tables

Abbreviations

EGFR	Epidermal growth factor receptor
IHC	Immunohistochemistry
NSCLCs	Non-small cell lung cancers
P. glaucum	*Pennisetum glaucum*
SCLCs	Small cell lung cancers
TKI	Tyrosine kinase inhibitor
WHO	World Health Organization

Chapter 1

Introduction

Lung cancer is a primary bronchogenic carcinoma. Numerous genetic alterations and family aggregation are characteristics of lung cancer. It is primarily classified as either non-small cell carcinoma or small cell carcinoma. The two types of lung cancer differ in their development and dissemination. Lung cancer is one of the deadly diseases that affect people's lives and health around the world [1]. Globally, over 1.8 million new instances of lung cancer are diagnosed each year [2]. Lung cancer can be treated with radiation therapy, surgery, chemotherapy, and targeted therapy [3].

Small cell lung cancers, which typically progress to central mediastinal tumors, are the most dedifferentiated tumors. Lung hormone cells are the source of SCLCs. SCLCs, which account for 10% to 15% of all lung cancers, are distinguished by their high level of aggressiveness, rapid spread into regional lymph nodes and submucosal lymphatic arteries, and almost universal absent bronchial invasion [3]. The three most prevalent forms of lung cancer that are included in the general category of non-small cell lung cancer (NSCLC) are adenocarcinoma, squamous cell carcinoma (SCC), and the large-cell carcinoma [4]. Out of all lung cancer, about 80 - 85% is NSCLC, and the tumor has been identified in an advanced stage in more than 65% of cases [5].

In this category, adenocarcinoma is very common, making up half of all occurrences. SCC (NSCLC type), was the most often diagnosed lung cancer prior to

adenocarcinoma. SCC mostly starts at the origin of the tracheobronchial tree, although there are more and more cases being found in the periphery of the lung. One subtype of NSCLC that is an excluding diagnosis is large cell carcinoma. It lacks differentiation and cannot be further categorized using electron microscopy or immunohistochemistry (IHC) [4].

Lung cancer has been found to contain a number of targetable genetic changes. KRAS, EGFR, BRAF, PI3K, MEK, and HER2 are among the proto-oncogenes that have activating mutations [3]. Cell survival, migration, metabolism, differentiation, and proliferation are all regulated by protein TK gene families. Transmembrane proteins known as receptor tyrosine kinases are known to have ligand-controlled TK activity. EGFR (HER1, ErbB1), ErbB2 (HER2), ErbB3 (HER3), and ErbB4 (HER4) are the four members of the EGFR family [6]. Notably, EGFR is essential for controlling apoptosis, normal cell division, and other biological processes. Tumor-associated EGFR mutations are present in about 35% of NSCLC patients in East Asia and 10% in US [3].

One of the most significant pest-resistant C4 cereal crops and drought-tolerant, pearl millet (*Pennisetum glaucum*) is grown for combustion, grain, and fodder, mostly in the tropics and sub-tropics of Africa south of the Sahara (Mali, Niger, Nigeria) and Asia (India). Millions of people in developing countries rely on millets as a staple diet because they are high in nutrients and bioactive phytochemicals like phenolic compounds, which may have health advantages. Given that millets are underutilized grasses in wealthy nations, an in-depth study of the chemical constitution and biological activity of millet grains needs to be conducted [7].

The use of computational methods in drug development is a new field that has the potential to greatly enhance the process of finding novel medications for various illnesses, including cancer. Molecular docking and molecular dynamic simulations are widely employed in computational drug development to identify novel compounds using in silico techniques [8, 9]. Phytocompounds are significant source of novel medicinal products. It is important need to develop more phytomedicines in comparison to traditional cancer treatments. Better binding affinities allow phytocompounds to prevent or impede the growth of tumors [10].

1.1 Problem Statement

Lung cancer, is a serious global health concern. NSCLC is the most common subtype, urging need for novel treatment approaches. Current chemotherapeutics exhibit significant limitations, thereby hindering effective cancer management.

1.2 Hypothesis

Novel therapeutic pharmaceuticals can be established from phytocompounds found in plants. In NSCLC, mutated EGFR can be inhibited by the phytochemicals from Pearl millet, thus leading to new therapeutic approach.

1.3 Aims and Objectives

To identify and characterize phytochemical-based drug candidates from *Pennisetum glaucum* for NSCLC treatment using computational approaches focused on EGFR inhibition and pharmacological profiling. The objectives are:

- To identify the phytochemicals of *Pennisetum glaucum* as inhibitors of EGFR protein.
- To assess the chemical interactions between chosen substances and important lung cancer protein targets.
- To rank and prioritize substances according to their anticipated inhibitory capabilities and binding affinities against lung cancer.

Chapter 2

Literature Review

2.1 Cancer

Cancer is characterized by aberrant cell development by several gene expression alterations, which unbalance cellular proliferation and apoptosis. As a result, the population of cells grows, invading neighboring tissues and spreading to other locations [11]. Cancer has been identified as leading cause of mortality in both developed and developing nations globally over the years, making it a chronic public health concern [12, 13].

Survival, self-sufficiency of growth signals, immortality, gene instability, avoiding programmed cell death, continual angiogenesis, unrestricted replication, and a variety of mutations are all unique characteristics of cancerous cells [14].

2.2 Cancer Treatment

Among the often employed methods for treating distinct cancer kinds are surgery, radiotherapy, and chemotherapy. Of these, radiotherapy and chemotherapy are being utilized more frequently to treat various metastatic cancers. Fatigue, hair loss, nausea, vomiting, anemia, diarrhea, skin difficulties, nerve problems, muscular abnormalities, bladder and kidney irritation, and numerous fertility and sexual

concerns are among the adverse effects that are common with the medications used for chemotherapy and radiation therapy [15]. Furthermore, commercially available anti-cancer medications have the potential to be harmful to the human body and cause a number of major side effects, including baldness, bone marrow depression, a decrease in WBC count, and a loss of immunity and self-confidence [16].

2.3 Lung Cancer

An out-of-control, malignant development that forms in the lung tissues and results in tumorous growths is known as lung cancer [17]. Millions of individuals are affected by lung cancer each year, making it the most widespread cancer globally and leading cause of cancer-related deaths [18]. Furthermore, it is anticipated that by 2035, the number of lung cancer deaths would rise to 3 million cases [17]. Lung cancer is responsible for more deaths overall than breast, colon and prostate cancer combined. More than half of patients, approximately 55%, had metastatic lung cancer upon diagnosis, which contributes to the poor prognosis for lung cancers [19]. According to WHO, it is the most prevalent kind of cancer in men, ranking first in terms of both mortality and incidence, whereas in women, it ranked third in respect of incidence and second in terms mortality [20].

2.4 Risk Factors

Tobacco use, radon exposure, air pollution from both indoors and outdoors, genetic mutations, and an imbalanced diet are some of the risk factors that can raise the chance of developing lung cancer [1]. Ionizing radiation, which is present in patients with a prior diagnosis of breast cancer or Hodgkin lymphoma, environmental pollutants including secondhand smoke, metals (arsenic, nickel, and chromium), radon, and polycyclic aromatic hydrocarbons are additional factors linked to an elevated risk of lung cancer. Furthermore, it has been determined that HIV infection, alcohol consumption, and a history of pulmonary fibrosis are risk factors for lung cancer [21].

2.5 Symptoms

Lung cancer frequently exhibits no signs in its early stages. Lung cancer symptoms usually indicate a benign condition because most of them are either non-specific or generic respiratory symptoms. People over 40 who have two or more of the following symptoms, or who have ever smoked and have one or more of the following symptoms, should be administered an urgent chest X-ray (should be performed within two weeks) to screen for lung cancer [22].

- Cough
- Fatigue
- Weight loss
- Chest pain
- Appetite loss
- Shortness of breath

2.6 Types of Lung Cancer

Lung cancer comes in two primary kinds:
1) SCLC (15%)
2) NSCLC (85% of patients).

2.6.1 SCLC

SCLC is a high-grade neuroendocrine carcinoma with a very bad prognosis that primarily affects smokers who are currently or were formerly smokers. SCLC represents 15% of lung cancer cases. With a lag time of roughly 30 years, the prevalence of SCLC—one of the malignancies with the highest epidemiological relation to tobacco—tends to resemble the prevalence of smoking [23–25].

2.6.2 NSCLC

NSCLC is divided into: large cell, squamous cell, and adeno carcinoma.

2.6.2.1 Large Cell Cancers

About 5% to 10% of all lung cancers are of this type, and its frequency is decreasing as a result of improved classification of more weakly differentiated adenocarcinomas and squamous cell carcinomas made possible by current immunophenotyping techniques. Large cells with lots of cytoplasm and big nucleoli make up these tumors, which are usually poorly differentiated [23–25]. Even when treated with certain combination therapies, patients with advanced NSCLC have a poor prognosis [3]. Therefore, it's critical to look into novel therapies that could improve NSCLC patients' outcomes [26]. The clinical data in Fig 2.1, highlights key insights into lung cancer cases, focusing on NSCLC (non-small cell lung cancer) and SCLC (small cell lung cancer). It presents their relative prevalence among all lung cancer cases, the distribution of NSCLC and SCLC subtypes, and the 5-year survival rates for lung cancer patients.

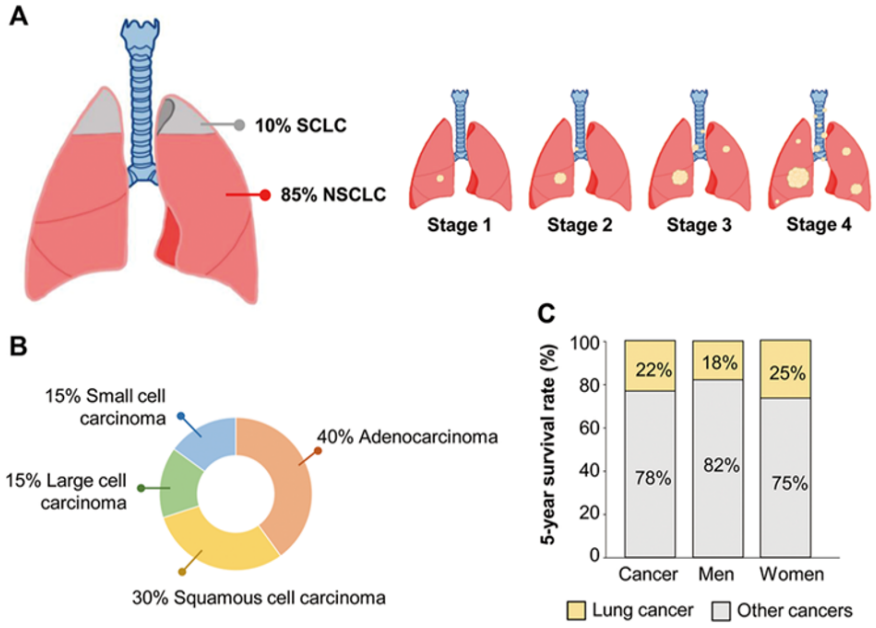


FIGURE 2.1: Clinical data for NSCLC patients. (A) SCLC and NSCLC as a percentage of all lung cancer cases. (B) The proportion of different types of NSCLC and SCLC. (C) 5-year survival rate for lung cancer patients [27].

2.6.2.2 Adenocarcinoma

It covers 40% of lung cancer cases. The WHO classified initial stages of lung cancer as invasive, minimally invasive, or adenocarcinoma in situ, depending on the degree of invasiveness; an adenocarcinoma with a lepidic pattern along with a diameter of less than 3 cm is called an adenocarcinoma in situ.

2.6.2.3 Squamous Cell Carcinomas

It accounts for 25-30% of lung malignancies and is usually in airway epithelium.

2.7 Treatment for NSCLC

2.7.1 Stage I and II NSCLC

The greatest option is surgery for long-term survival, which is primarily treating operable and removable early stage illness (Stage I and II). For individuals with stage I NSCLC, the 5 year survival rate following surgical resection is 60%–80%, but for those with stage II NSCLC, it is 30%–50%.

2.7.2 Stage III NSCLC

Over 70% of patients with NSCLC get a diagnosis in stage III or stage IV, which is considered advanced illness. The location of the tumor and whether it is removable dictates treatment strategies for stage III NSCLC, like chemotherapy, radiation therapy, and surgical resection.

2.7.3 Stage IV NSCLC

At 40% of newly diagnosed NSCLC patients, stage IV NSCLC patients are treated based on a variety of factors, such as histology, comorbidity, molecular genetic characteristic of the cancer.

The standard treatment options for stage IV NSCLC disease include combination chemotherapy, the internal endoscopic radiation therapy, targeted therapies, and any laser therapy as needed. Surgery may also be used in certain cases to relieve symptoms related to the disease, similar to radiation therapy.

2.7.4 Recurrent NSCLC

An NSCLC that has reappeared or progressed after initial chemotherapy, radiation therapy, or surgery is referred to as recurrent or relapsed [28].

2.8 Growth Factor Receptors

Proteins known as growth factor receptors (GFRs), which are activated by interacting with their ligands, have important roles in tumor growth, angiogenesis, metastasis, cell survival, migration, cell death, neovascularization, differentiation, organogenesis, and chemoresistance [29, 30].

Single-transmembrane proteins known as growth factor receptors may contain enzymatic (kinase or phosphatase) activity may be linked to intracellular enzymes [31]. The role of epidermal growth factor (EGF) and its specific receptor, EGFR, in oncogenesis is very important [32, 33]. EGFR is a member of the receptor tyrosine kinase family, which also includes EGFR/ERBB1, HER4/ERBB4, HER2/ERBB2/NEU and HER3/ERBB3. 15% of NSCLC patients in Western population and 35% of those in the Asian population had EGFR mutations [34].

Brain, Bladder, breast, colon, digestive, lung, thyroid, prostate, kidney, and ovarian cancer cells were among the many tumors that contained VEGFR1-3 genes; breast, gastric, endometrial, and lung carcinomas contained FGFR-2 genes; bladder, breast, lung, and glial carcinomas overexpressed HER-2 genes; aggressive metastatic ovarian, gastric, lung, breast, invasive urothelial bladder, and endometrial carcinomas contained the HER-3 gene [35, 36]. These receptors are a crucial target for effective cancer treatment due to their role in the genesis of cancer.

2.9 EGFR

The epidermal growth factor receptor, or erbB-1, is member of tyrosine Kinase receptor's erbB family. Numerous ligands have the ability to attach to and activate EGFR binding. Ligands create dimerizations when they attach to receptors. The intracellular tyrosine kinase action area of EGFR then undergoes autophosphorylation, which creates a binding site for the signaling molecule and initiates the downstream signaling pathway, so facilitating cell survival, differentiation, and proliferation (Fig 2.2).

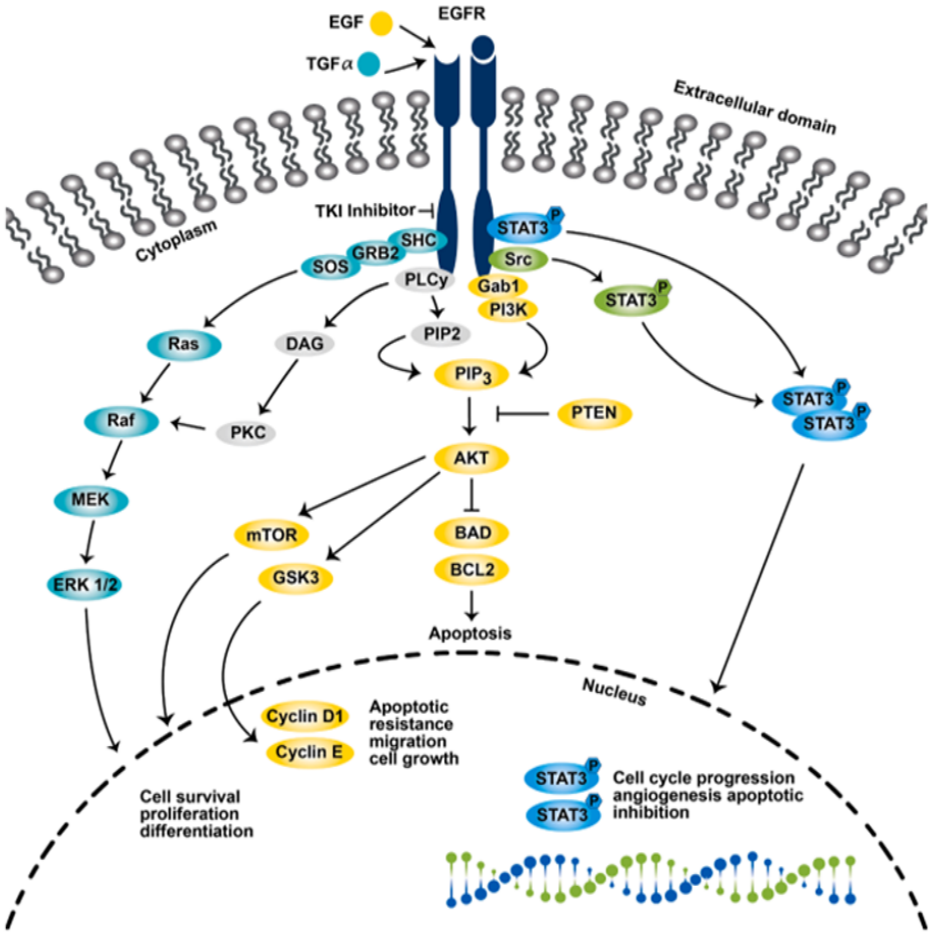


FIGURE 2.2: EGFR signaling [37].

It plays important role in number of solid tumors, the most common of which being NSCLC. EGFR inhibitors, which prevent their function, are also created as a result of the mechanism of action of EGFR and the identification of carcinogenic factors [38].

Mutations, amplifications and deletions in sequences encoding the TK domain can all activate EGFR [39]. The PI3K/Akt, Ras/MAPK, Src and phospholipase C (PLC), pathways are among the canonical cellular signaling pathways that are activated when EGF binds to the extracellular domain of EGFR, causing the receptor to dimerize, autophosphorylate, and activate [40].

2.10 EGFR Inhibitors

Monoclonal antibodies and tyrosine kinase inhibitors (TKIs) are used for EGFR targeting. EGFR TKIs are composed of first generation gefitinib & erlotinib, the second generation dacomitinib & afatinib, & third generation osimertinib [21].

2.11 EGFR-TKIs

Patients with advanced localized or metastatic NSCLC who have EGFR-TKI-sensitizing mutations are treated with first-generation or second-generation EGFR-TKIs, such as erlotinib, gefitinib or afatinib [41, 42]. Osimertinib, an orally, third-generation, irreversible EGFR-TKI, preferentially inhibits both EGFR T790M resistant mutations and EGFR-TKI-sensitizing, but has minimal impact on wild-type EGFR [43, 44].

2.12 Phytochemicals

Plant compounds known as phytochemicals have defensive qualities that give them an edge when it comes to disease infection; they do not react. This easily obtainable nutrient, which is found in large quantities in plants, contains compounds that give plants their physiological or organoleptic characteristics, including color. In addition to protecting plants, these phytonutrients may also shield people from conditions including cancer, heart disease, metabolic syndrome, and others. Many plant metabolic activities generate molecules known as secondary metabolites [45].

Phytochemicals are naturally occurring substances derived from plants that are essential for development of new medications and for treatment of cancer. Podophyllotoxin analogs, vinca alkaloids like vinblastine and vincristine, and taxol analogs are a few common examples. Often, these phytochemicals function by regulating molecular pathways connected to the onset and metastasis of cancer. Some of the specific methods include boosting antioxidant status, deactivating carcinogens, stopping proliferation, triggering an arrest in cell cycle and cell death, and immune system regulation [12].

2.13 Pearl Millet

Millets are becoming more and more well-known for their excellent nutritional value and tolerance to the effects of climate change. The word "millet" refers to a wide variety of species from different genera, the most common of which are proso, foxtail, finger, and pearl millet. Pearl millet (Fig 2.3) is valuable for stover as a source of dry fodder, but it is mostly grown for grain production in Sub-Saharan Africa and Asia. It is generally grown as a fodder crop in other nations. A cereal crop that can withstand heat and drought, pearl millet produces a consistent output of grain and fodder on sandy, poor soils in hot, dry climates [46].

FIGURE 2.3: Pearl millet crop [47].

2.14 Phytochemicals in Pearl Millet

The presence of active substances such flavonoids, saponins, tannins, alkaloids, polyphenols, and anthraquinones has demonstrated that plants in the genus Pennisetum are active in terms of phytochemicals and pharmacological activities [48].

Phytochemicals including terpenoids, phenols, glycosides, flavonoids and steroids were found in *P. glaucum* [49].

Pearl millet contains a wider variety of phenolic compounds; according to this study, *P. glaucum* oil contains seven different compounds, including three major acids that are phenolic (trans-cinnamic, hydroxybenzoic acids and protocatechuic), four minor acids (gentisic, p-Coumaric gallic, and caffeic), and two flavonoids (quercetin and catechin) [50]. When phenolic acid concentrations were compared to those of other cereals, pearl millet had higher levels (64.8 mg/kg) than sorghum (27.3 mg/kg) [51]. Tricine, 3, 4 Di-OMe luteolin, acacetin, and 4-OMe tricin were the flavonoids that were most frequently found. Vanillic acid, syringic acid, para-hydroxybenzoic acid, melilotic acid (2-hydroxy benzene propanoic acid), and salicylic acid are the five phenolic acids that have been identified [52]. Gallic acid, p-coumaric acid, chlorogenic acid, syringic, ferulic acid, hydroxycinnamic acid, ellagic acid, quercetin, and apigenin are among the polyphenols found in pearl millet [53].

2.15 Taxonomic Hierarchy

The taxonomic hierarchy of *Pennisetum glaucum* is given in table 2.1 [54].

TABLE 2.1: Taxonomy of *P. glaucum* [54].

Domain	Eukarya
Kingdom	Plantae
Sub-kingdom	Tracheobionta
Tribe	Paniceae
Division	Magnoliophyta
Subdivision	Spermatophyta
Class	Liliopsida-monocotyledons
Sub-class	Commelinidae
Order	Poales
Family	Poaceae
Subfamily	Panicoideae
Genus	*Pennisetum*
Species	*glaucum*

2.16 Natural Compounds as Inhibitors of EGFR Protein

Natural compounds are an important source of pharmaceuticals and are used in the research and discovery of novel medications. Because of their distinct chemical diversity and biological activity, natural chemicals have been a successful source of new medications in recent decades. Furthermore, due to their biological actions, natural chemicals hold a prominent place in anti-tumor therapy. Because they target the EGFR pathways, natural chemicals are very important in the discovery and development of anti-tumor medicines. Through EGFR signaling pathways, phytochemicals from several chemical categories have been shown to have anti-cancer properties. Since TK is the primary driver of tumors, its aberrant activity has emerged as a major target for clinical therapies. Finding and categorizing the chemicals from natural sources that are EGFR-TKIs is therefore crucial [55].

2.17 Inhibitors against EGFR Protein of in *P. glaucum*

Millet contains plant protease inhibitors, dietary fibers, and bioactive peptides that have the ability to suppress cell proliferation, cause apoptosis, and interact with the gut microbiota to potentially have anti-cancer effects. High levels of phytochemical activity can be seen in finger millet, tiny millet, foxtail millet, pearl millet, and kodo millet.

Among the varieties, pearl millet contains the highest concentration of antioxidant components, indicating its exceptional antioxidant capacity. The methanol extracts of minor millets have been discovered to contain the majority of possible phytochemicals, such as flavonoids, xanthophylls, sterols, phenols, xanthine, terpenoids, and carotenoids. Furthermore, it has been noted that phytochemicals have a number of health advantages, including antihyperlipidemic, anticancer, antidiabetic, and antioxidant properties [56].

2.18 Molecular Docking

The "molecular docking" predicts the preferred orientation of the ligand toward the receptor (a protein) in order to produce a stable molecule. By using scoring functions, preferred orientation can be utilized to predict the level of ligand-protein connection or binding affinity. Docking is widely used to determine the direction in which drug candidates might attach to protein targets in order to predict the affinity and activity associated with a treatment. For this reason, docking is crucial to the drug design and discovery process. The goal of molecular docking is to achieve an optimal conformation and computationally model the molecular identification process in order to reduce the system's free energy. Finding a new medication is a really challenging process. The primary method used in modern drug development is the in-silico approach. The application, adoption, and appreciation of computer-aided approaches in the drug research and development process are growing quickly [57].

A most popular method for forecasting ligand binding affinity and mechanism of action is protein–ligand docking. One effective method for CADD is protein-ligand docking. CB-Dock is a user-friendly blind docking web server. It simply needs a protein file in the PDB format, and the ligand file in exact SDF formats to automatically predict binding modes without knowing the binding sites. The sizes and centers of all of the N cavities are then determined by CB-Dock, which also predicts the protein's cavities. For docking, AutoDock Vina retrieves the pdbqt files of any size, via any center [58].

Chapter 3

Research Methodology

For the identification of compound against NSCLC from *P. glaucum*, methodology flowchart is given in Fig 3.1.

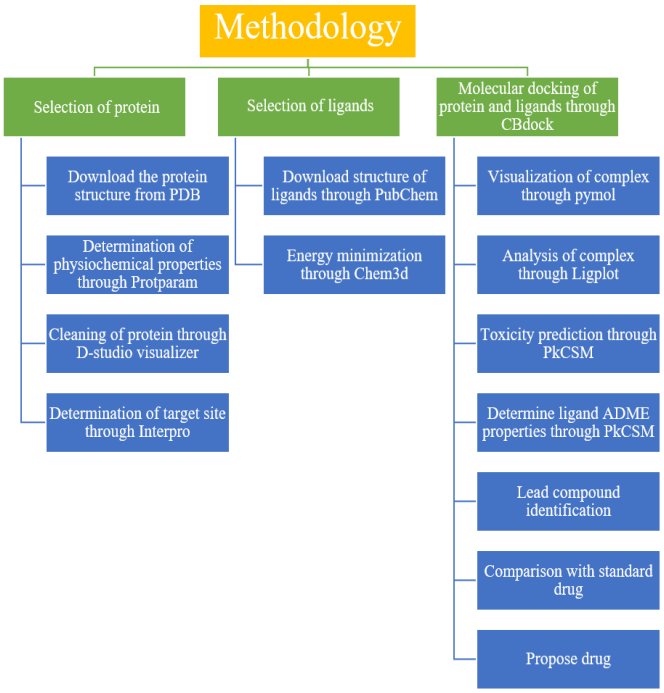


FIGURE 3.1: Methodology flowchart of research.

3.1 Selection of Disease

One of the major causes of mortality is cancer. NSCLC is recognized as the most persistent type of cancer worldwide among all cancer forms. Radiation treatment,

chemotherapy, and surgical excision are some of the medical procedures used in the traditional care of lung cancer.

This sort of treatment, however, damages adjacent normal cells and is not particularly effective. This unfilled medical necessity has increased the need for novel therapeutic strategies.

3.2 Protein Selection and Preparation

3.2.1 Target Protein Selection

First of all, targeted protein was selected i.e EGFR from literature review. EGFR gene become mutated in NSCLCs and leads to over activation of EGFR protein. This type of mutation accounts for approximately 80-85% of all lung cancers.

3.2.2 3D Structure of Protein

A 3D structure of EGFR protein was acquired from Protein data bank (PDB) in pdb format. The PDB provided the crystal structure of the EGFR protein.

3.2.3 Physiochemical Properties of Protein

The physical and chemical attributes, such as molecular weight, theoretical pI, amino acid composition, atomic composition, aiphatic index, instability index, estimated half-life, extinction coefficient, and grand average of hydropathy of the EGFR protein, was computed using Expasy ProtParam tool [59].

3.2.4 Cleaning of Protein

Cleaning of targeted protein involves removal of water molecules and ligands. And this purification was done using D-studio visualizer.

3.2.5 Functional Domain

Protein functional domains and conserved sites are identified via the InterPro database. It has classified proteins into families based on their sequences [60]. The EGFR protein's functional domains was identified using InterPro database.

3.3 Selection of Ligands

Ligands of *Pennisetum glaucum* was selected from literature review.

3.3.1 Structures of Ligands

A public database of chemical structures and the findings of their biological analyses is available at PubChem [61]. In PubChem, a vast range of chemical information is accessible, including syntheses, journal articles, patents, drug labeling, spectrum information, clinical studies, and molecular structures and characteristics [62]. From PubChem, structures of ligands were acquired in sdf format.

3.3.2 Energy Minimization of Ligands

With the downloaded ligands, energy reduction was carried out prior to molecular docking via Chem3d Pro 12.0. The conformational structure with the lowest total potential energy is the result of the energy minimization method [63].

3.4 Molecular Docking

Molecular docking is a computational approach (insilico) that selects the best posture each molecule produces to a rank-order. This is done by determining the correct binding pose of a target protein and ligand complex and evaluation of the strength of the interaction using a range of scoring functions.

Docking techniques aim to fit a ligand into a protein's binding site by optimizing hydrophobic, electrostatic, and steric interactions, ultimately predicting the binding free energy of the target protein [64].

A well-known docking application named Autodock The user-friendly blind docking web service CB-Dock, which predicts a protein's binding sites and determines its centers and sizes using a special curvature-based cavity identification method, is used to perform docking with Vina. This makes it possible to forecast binding modes automatically without requiring knowledge of binding sites. For our chosen ligands against the targeted protein, we employed CB-Dock in this instance [58].

3.5 Ligand Protein Interaction

The LigPlot tool is used to determine intermolecular interactions. LigPlot is useful for this purpose since it displays every detail of the interactions in addition to producing a visual depiction of them. LigPlot was recently re-released as LigPlot+ after being upgraded [65]. The ligand and receptor interactions were analyzed by opening the pdb file of the optimum docked complex in LIGPLOT+ [66].

3.6 Virtual Screening

3.6.1 Lipinski Rule of Five

Lipinski's Rule of Five (RO5) physicochemical parameter standards would be met by an ideal therapeutic molecule. It forecasts the drug-like characteristics of a chemical molecule intended for oral administration that has a certain biological function. A compound will have greater bioavailability and improved pharmacokinetic characteristics in the organism's metabolic process if it fulfills these criteria [67]. Following are the Lipinski's rule of five:

1. Hydrogen bond acceptors ≤ 10

2. Hydrogen bond donors ≤ 5
3. Molecular weight < 500 g/mol.
4. LogP < 5
5. Rotatable bonds < 10

Calculation of Lipinski's Ro5 of compounds were done using the pkCSM tool.

3.6.2 ADMET Properties

In addition to biological activity, a successful drug's therapeutic effect depends on a suitable ADMET profile [64]. Pre-clinical optimization, which addresses the physicochemical properties and in silico toxicity assessment, is the next phase in the computer-aided drug design pipeline that must be completed before a hit molecule can be generated and deployed to use. A compound's ADME characteristics are determined by its pharmacokinetic profile. The method used for anticipating and optimizing pharmacokinetic and toxicity qualities is called pkCSM [68].

Utilizing the pkCSM server, the ADMET characteristic of particular phytochemicals were predicted. To anticipate the ADMET properties, SMILES of the chosen compounds was uploaded to the pkCSM server. The pkCSM server provided a range of parameters for the profiles of selected compounds, including those related to absorption, distribution, metabolism, excretion and toxicity [69].

3.6.3 Lead Compound Identification

By virtual screening through Lipinski RO5 and ADMET analysis lead compound was identified. After that, analysis of docking scores, bonding amino acids, protein ligand interactions were then carried out. Lead compounds are the ones which satisfied all these requirements.

3.7 Standard Drug Selection and Screening

3.7.1 Standard Drug Selection

Approved drug for mutated EGFR in NSCLC was identified through KEGG database. A database called KEGG DISEASE connects disease genes, pathways, medicines, and diagnostic markers. KEGG DRUG comprises all licensed drugs in the US and Japan [70].

3.7.2 Standard Drug Screening

The physiochemical properties, 3D structure, toxicity and SMILES were obtained from PubChem and ADME properties of standard drug was identified by pkCSM.

3.7.3 Drug Docking

The selected standard drug was docked with EGFR protein using CB-dock.

3.8 Comparison of Lead Compound with Standard Drug

The ADMET properties, physiochemical properties, and docking results, of standard drug and lead compound were compared.

Chapter 4

Results and Discussion

4.1 Selection and Preparation of Protein

4.1.1 Structure of Protein

Different experimental techniques exist to determine the structure of target protein. These include X-ray crystallography, cryogenic electron microscopy or nuclear magnetic resonance (NMR). These structures can be accessed from Protein data bank (PDB) [71]. The protein data bank (PDB) provided the 3D structure of the EGFR protein (Fig 4.1), accessed with accession code 1XKK, DOI https://doi.org/10.2210/pdb1XKK/pdb.

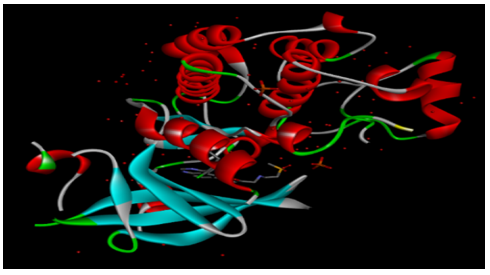


FIGURE 4.1: 3D structure of the Tau protein.

4.1.2 Protein Purification

The EGFR protein crystal structure acquired from PDB (accession No: 1XKK) was refined to prepare it for docking (Fig 4.2). For this purpose, Biovia Discovery

Studio Visualizer was used to remove heteroatoms and water molecules. After purification EGFR structure was saved in PDB format [72].

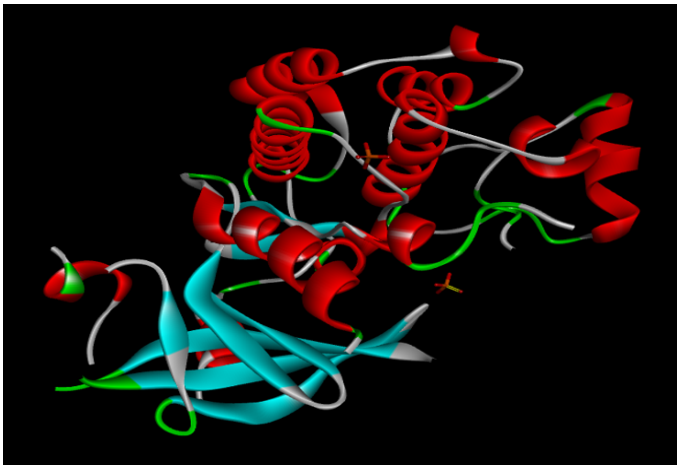


FIGURE 4.2: Purified structure of EGFR.

4.1.3 Determination of Physiochemical Properties of Protein

The chemical and physical properties of target protein sequence, were analyzed using ExPASy's ProtParam tool. These properties included molecular weight, extinction coefficient, theoretical pI, atomic composition, instability index, estimated half-life, amino acid composition, total No. of residues with positive charge (Arg + Lys), total No. of residues with negative charge (Asp + Glu), grand average of hydropathicity (GRAVY) and aliphatic index [73].

A pH value at which molecule carried no electric charge is designated as isoelectric point (pI). This notion finds its significance in zwitterionic molecules (like proteins, peptides and amino acids). This value indicates the global basic or acidic nature of zwitterionic molecule. Compounds having pI > 7 are basic, and the ones with pI < 7 are acidic [74]. In order to predict stability of protein in vivo, a primary structure dependent method available is dipeptide composition based Instability Index (II). According to this method proteins are stable if they have II value below 40 [75]. Aliphatic index (AI) is the relative vol. of protein's aliphatic side chains. This serves a crucial role in thermal stability of protein. Higher the Aliphatic index, more thermally stable the protein is [76].

The average hydropathy value of peptide/ protein is represented by GRAVY (Grand Average of Hydropathy) [77]. Hydropathy index of amino acid shows whether the side chains are hydrophobic or hydrophilic. This idea was first proposed by Jack Kyte and Russell F. Doolittle in 1982. Greater the No., more hydrophobic amino acid is. Isoleucine (4.5) and valine (4.2) are the amino acids with the highest hydrophobicity. On the other hand, arginine (-4.5) and lysine (-3.9) are the amino acids with the highest hydrophilicity. This is very significant when considering the 3D structure of protein as hydrophilic amino acids are commonly present on surface, while hydrophobic ones are buried inside [78].

The xtincttion coefficient of protein is usually accurate at wavelength of 280nm. In order to determine this, a common approach used is amino acid sequence based theoretical prediction. These theoretical predictions are made based on molar UV absorption (280nm) of tryptophan, tyrosine, and cystine residues [79]. Lys and Arg are cationic with positive charge while glutamic acid and aspartic acid are anionic with negative charge [80]. Table 4.1 shows that protein EGFR was acidic in nature and it was unstable. It has low GRAVY which indicates better interaction with water molecules. A higher aliphatic index indicates that a protein is more thermostable.

TABLE 4.1: Physicochemical properties of EGFR protein acquired via Protparam tool.

Parameters	**Values**
No. of amino acids	1210
Molecular Weight	134277.40
Theoretical pI	6.26
Total anion residues (Asp + Glu)	138
Total cation residues (Arg + Lys)	124
Instability Index (II)	44.59
Aliphatic Index	80.74
GRAVY	-0.316
Extinction Coefficient 1	128890
Extinction Coefficient 2	125140

4.1.4 Identification of Protein Functional Domains

The globular protein's smallest functional unit is protein domain. A plethora of structural domains have been identified. Every one of them is responsible for a particular function in cell. The protein-protein interactions is mediated by many such domains, by binding to ligand molecules' structural features or motifs [81]. The InterPro database (http://www.ebi.ac.uk/interpro/) brings together various predictive models, called signatures, which represent different protein domains, families, and functional sites. It combines data from many sources to give a more complete picture of how proteins work and their roles in organisms [82]. The UniProtKB database freely accessible at www.uniprot.org, is the most comprehensive high quality resource for sequence of proteins and their respective functional information. It gives a summary of computationally predicted or experimentally verified information. This information is provided by biocuration team [83]. From Uniprotkb, fasta format sequence of EGFR protein was obtained and then used in Interpro. The work ID for determining EGFR functional domain is accessible from https://www.ebi.ac.uk:443/interpro//result/InterProScan/iprscan5-R20240820-061233-0395-39757278-p1m/. Fig 4.3 shows that protein EGFR had 6 functional domains and chain with 1210 residues.

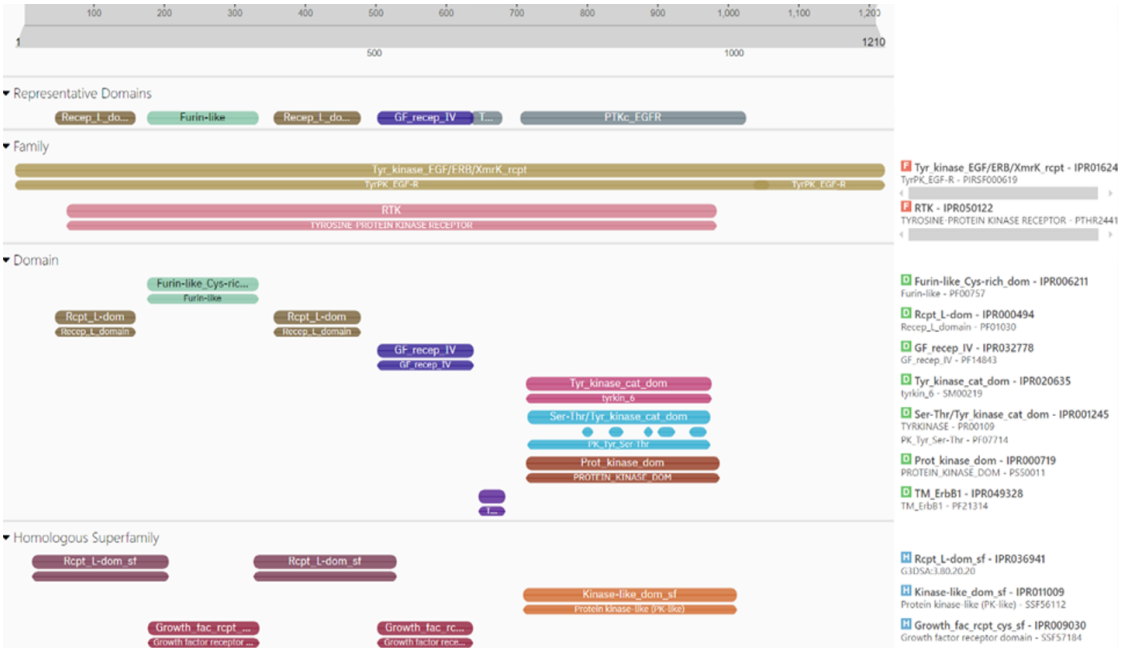


FIGURE 4.3: Functional domains of EGFR.

4.2 Ligands Selection and Preparation

An open chemistry database called PubChem provided information about chemical properties, structures, and biological activities. It is managed by National Institutes of Health (NIH). It mainly focuses on small molecules but also have macromolecules like peptides, lipids, carbohydrate, nucleotides etc. [71]. Literature review revealed 15 compounds, and their respective structures were procured from PubChem in SDF format. Table 4.2 shows molecular weight, formula and structure of all ligands obtained from PubChem. Using PyMol they were converted to PDB format. This was visualized in Discovery studio visualizer. Energy minimization was done by Chem3D and then file was saved as PDB format.

TABLE 4.2: Molecular weight, formula and structure of all ligands obtained from PubChem.

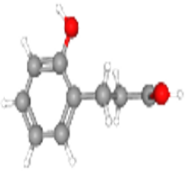
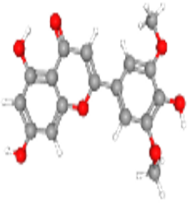
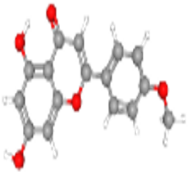
Sr. No.	Ligands	CID of PubChem	Mol. Formula	Mol. Wt. (g/mol)	3D Structure
1	Melilotic acid	873	C9H10O3	166.17	
2	Tricin	5281702	C17H14O7	330.29	
3	Acacetin	5280442	C16H12O5	284.26	
4	Gallic acid	370	C7H6O5	170.12	

Table 4.2: (Continued).

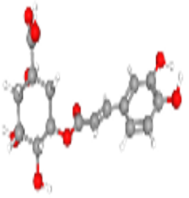
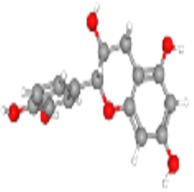
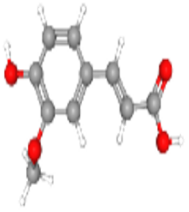
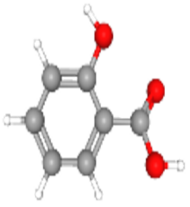
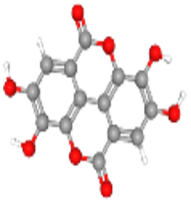
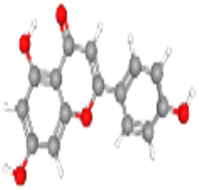
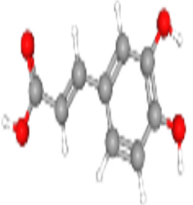
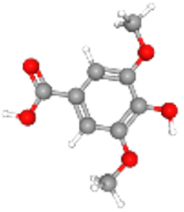
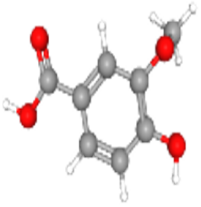
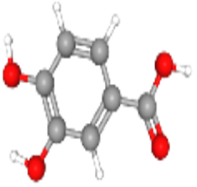
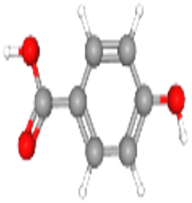
Sr. No.	**Ligands**	**CID of PubChem**	**Mol. Formula**	**Mol. Wt. (g/mol)**	**3D Structure**
5	Chlorogenic acid	1794427	C16H18O9	354.31	
6	Catechin	9064	C15H14O6	290.27	
7	Ferulic acid	445858	C10H10O4	194.18	
8	Salicylic acid	338	C7H6O3	138.12	
9	Ellagic acid	5281855	C14H6O8	302.19	
10	Apigenin	5280443	C15H10O5	270.24	
11	Caffeic acid	689043	C9H8O4	180.16	

Table 4.2: (Continued).

Sr. No.	Ligands	CID of PubChem	Mol. Formula	Mol. Wt. (g/mol)	3D Structure
12	Syringic acid	10742	C9H10O5	198.17	
13	Vanillic acid	8468	C8H8O4	168.15	
14	Protocatechuic acid	72	C7H6O4	154.12	
15	p-Hydroxybenzoic acid	135	C7H6O3	138.12	

4.3 Molecular Docking

The very easy and common computational structure based drug design (SBDD) approach is molecular docking (MD). This is being widely used since 1980s. This is preferred to be used when 3D structure of protein is available. The immense rise in availability and power of computers has aided popularity of molecular docking, as they increased No. and accessibility to small molecules and protein structures. The objective of molecular docking is to understand and predict molecular recognition. This is done at structural level by identifying potential binding styles, and energy level by prediction of affinity of binding.

Originally MD was developed for use on small molecules (lignads) and target large molecules (protein). There is however, growing interest in docking proteins with proteins, nucleic acid (DNA and RNA) with ligands, and nucleic acid, protein and ligands [71]. Table 4.3 shows docking scores of all ligands with highest docking score of Apigenin which was -8.4.

TABLE 4.3: Docking scores of 15 ligands with EGFR protein.

Sr. No.	**Compounds**	**Vina Score**	**Cavity Size**	**Center**	**Size**
1	Acacetin	-8.1	2263	37	32
2	Syringic acid	-5.3	407	45	18
3	Salicylic acid	-6.0	2263	37	32
4	Tricin	-7.8	2263	37	32
5	Melilotic acid	-6.0	2263	37	32
6	Catechin	-8.3	2263	37	32
7	Chlorogenic acid	-8.0	2263	37	32
8	Apigenin	-8.4	2263	37	32
9	Ferulic acid	-6.1	2263	37	32
10	Ellagic acid	-7.9	2263	37	32
11	Caffeic acid	-6.5	2263	37	32
12	p-Hydroxybenzoic acid	-5.9	2263	37	32
13	Gallic acid	-6.3	2263	37	32
14	Protocatechuic acid	-5.9	2263	37	32
15	Vanillic acid	-5.5	2263	37	32

4.4 Interactions of Ligands with Protein

A 2D schematic illustration is automatically generated by LIGPLOT program from PDB file input. The output file (coloured or black and white) gives simple and pictorial information about intermolecular interactions between ligands and proteins. It also shows the strength of interactions, H-bonds, hydrophobic inter-

actions and availability of atoms. The software is quite generic and can be used for any ligand. It can also show what kind of interactions occur between protein and nucleic acids [84].

Interactions of ligands with EGFR protein can be noted in Figure 4.4 – Figure 4.18 and Table 4.4. Acacetin and apigenin made 2 HBs, 13 and 11 hydrophobic interactions respectively. Caffeic, salicylic and p-hydroxybenzoic acid made 3 HBs, 10, 7 and 7 hydrophobic interactions respectively. Chlorogenic and syringic made 7 HBs, 10 and 2 hydrophobic interactions respectively. Protocatechuic, ellagic and ferulic acid made 2 HBs, 8, 8 and 7 hydrophobic interactions respectively. Gallic and melilotic made 4 HBs, 4 and 6 hydrophobic interactions respectively. Tricin and vanillic made 4 HBs, 10, 9 hydrophobic interactions respectively. Catechin made 6 HBs and 9 hydrophobic interactions.

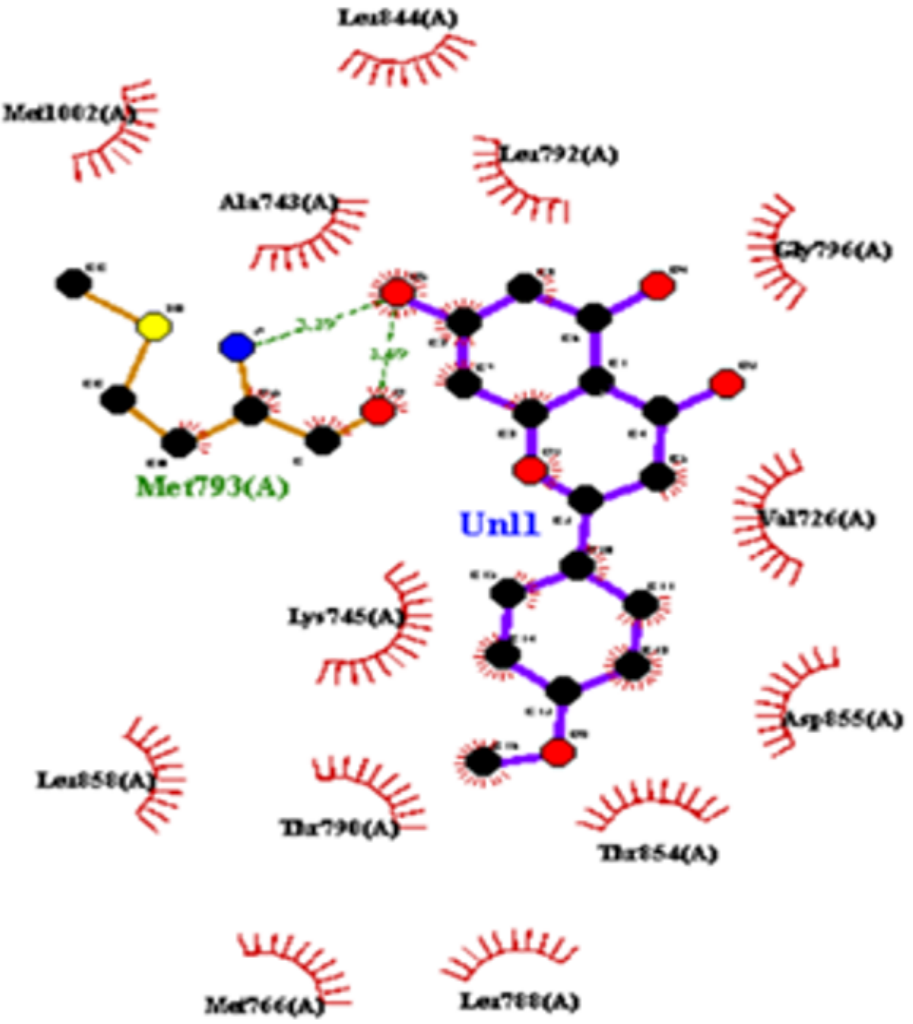


FIGURE 4.4: Docking of acacetin ligand with EGFR protein.

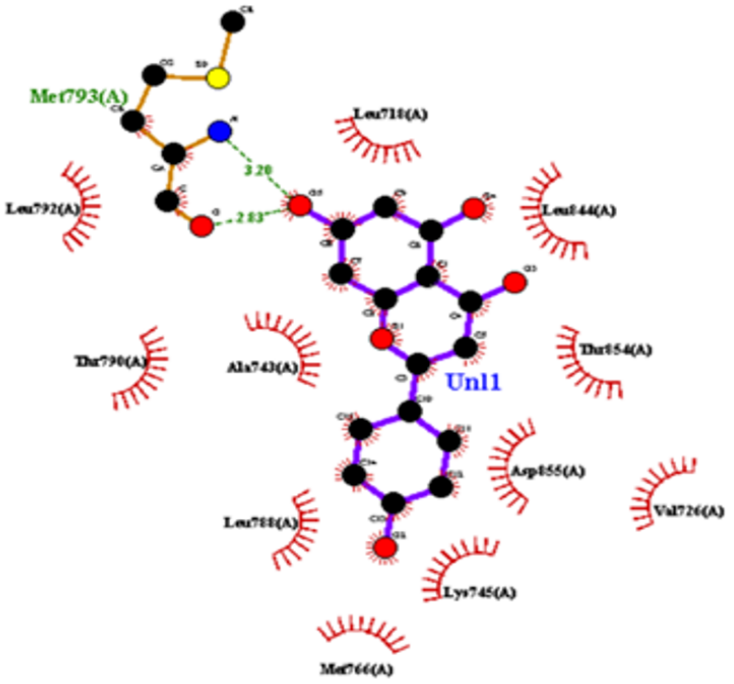


FIGURE 4.5: Docking of apigenin ligand with EGFR protein.

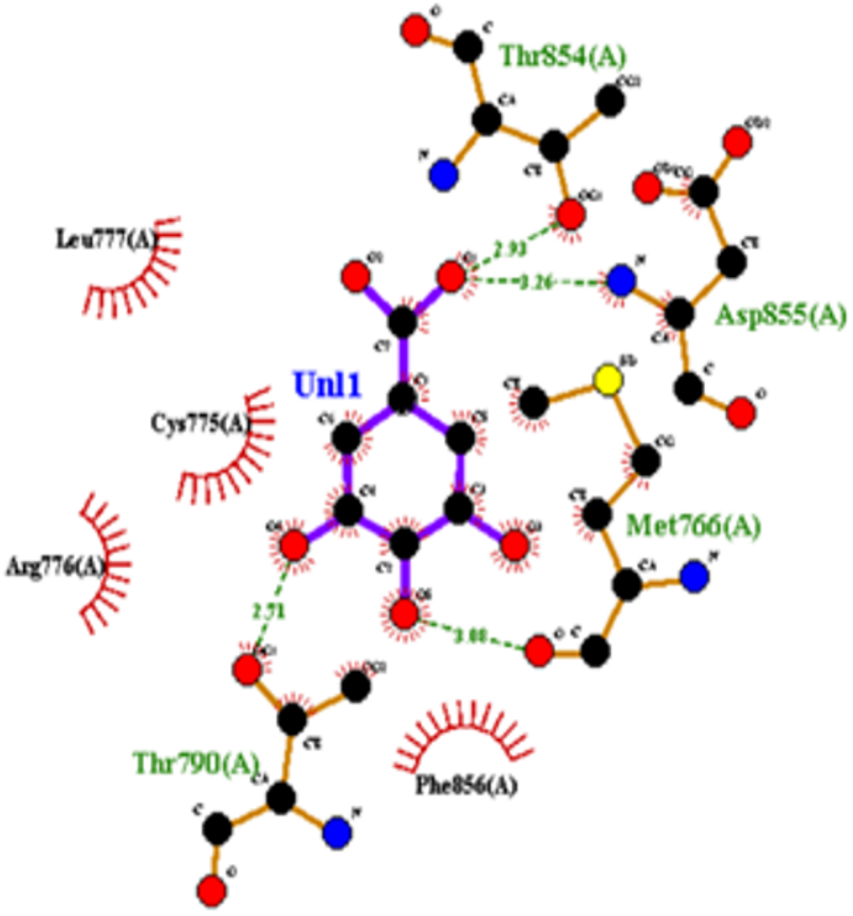


FIGURE 4.6: Docking of gallic ligand with EGFR protein.

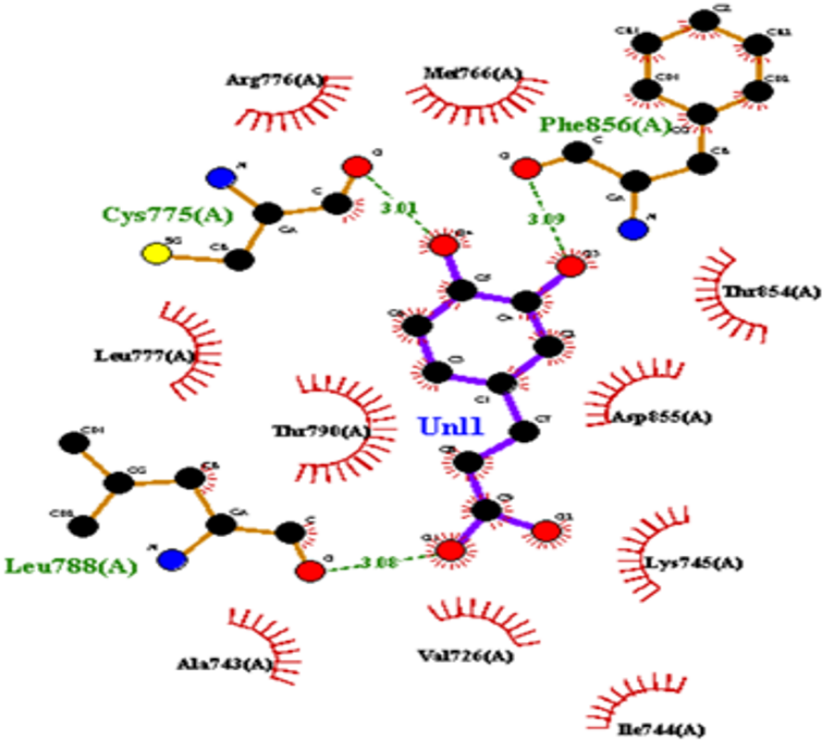


FIGURE 4.7: Docking of caffeic ligand with EGFR protein.

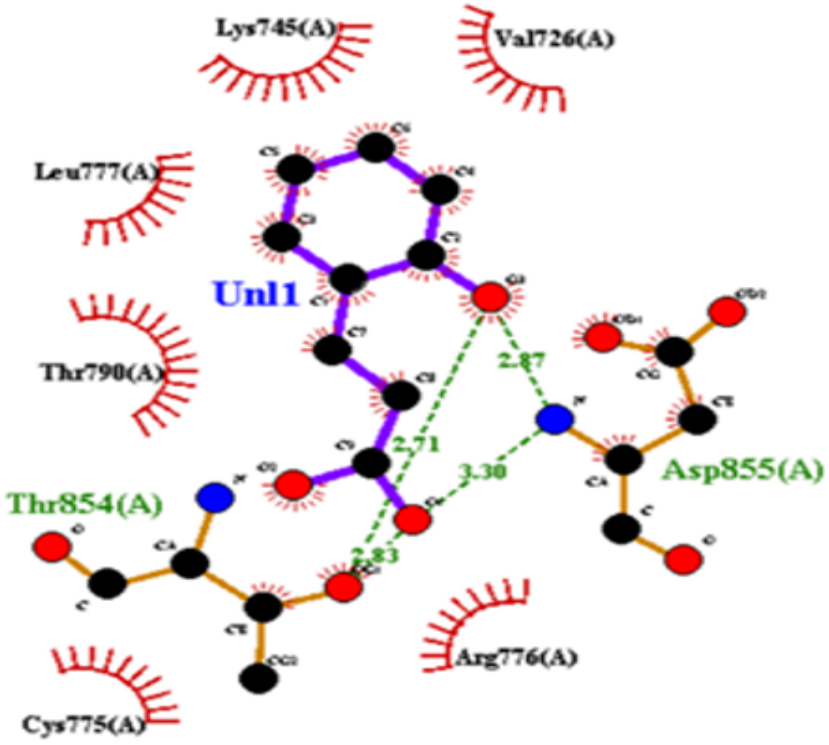


FIGURE 4.8: Docking of melilotic ligand with EGFR protein.

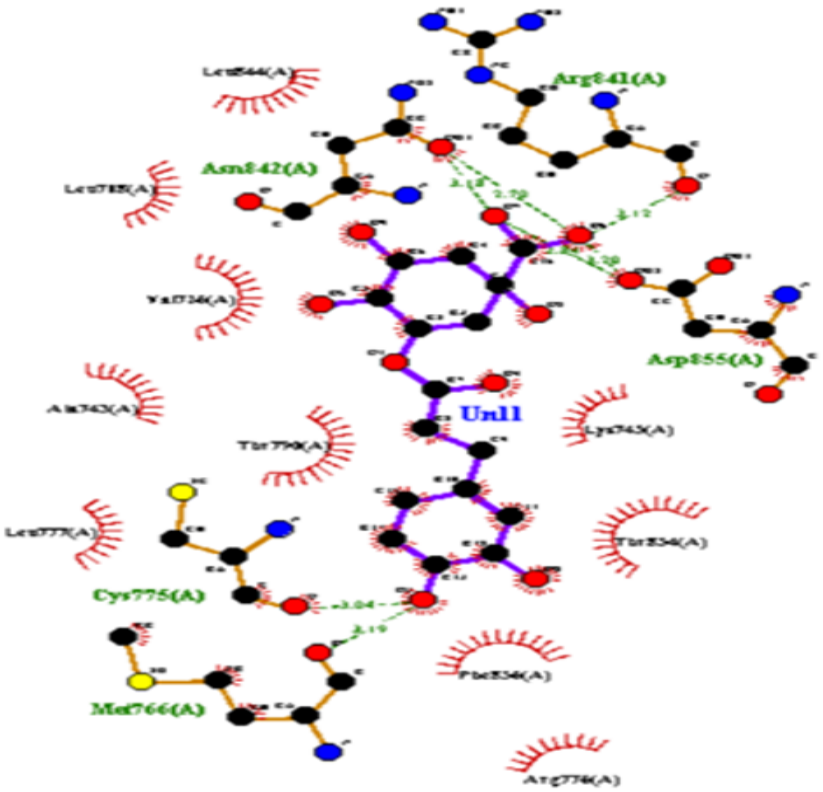


FIGURE 4.9: Docking of chlorogenic ligand with EGFR protein.

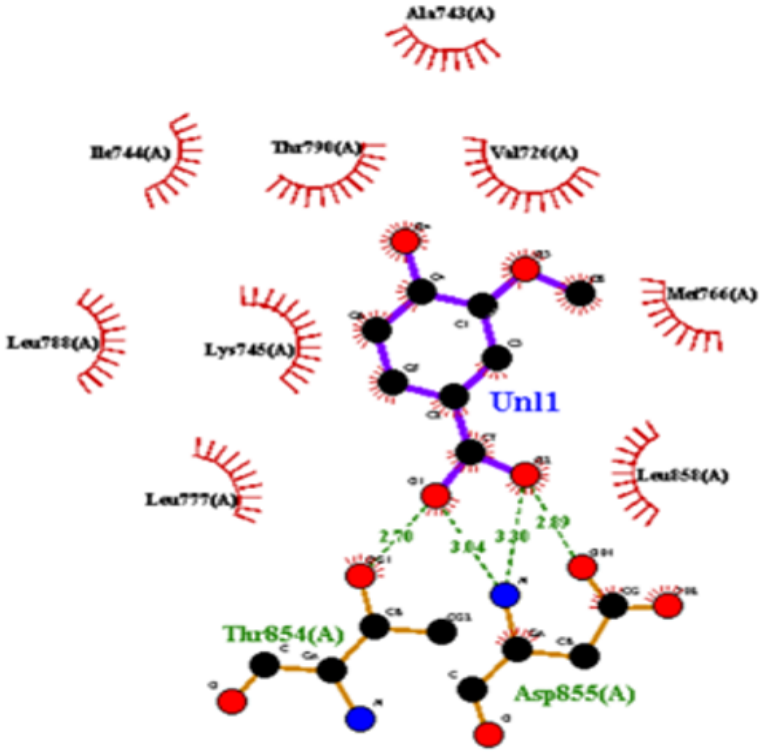


FIGURE 4.10: Docking of vanillic ligand with EGFR protein.

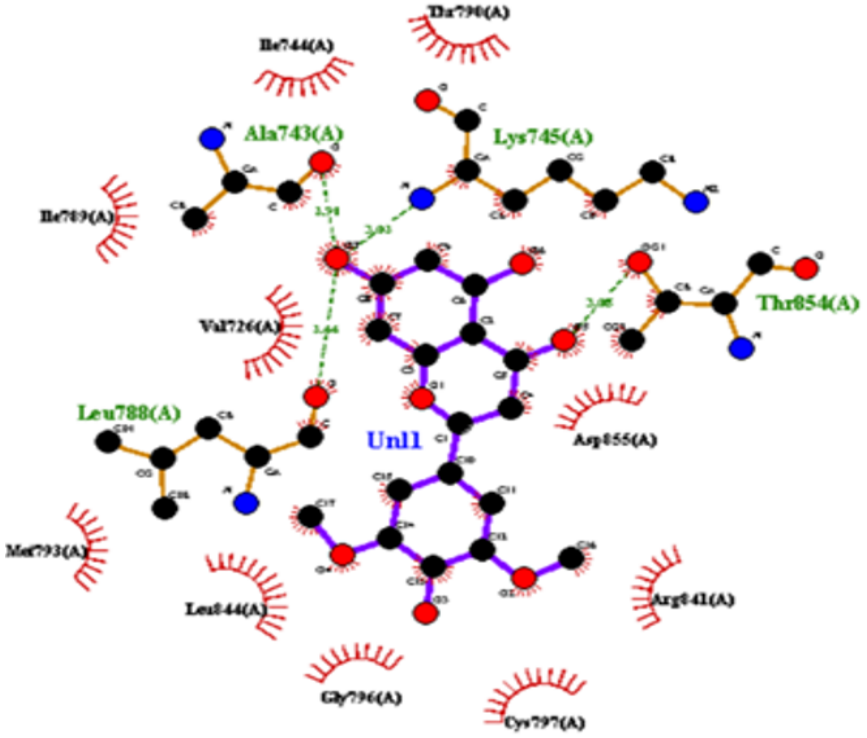


FIGURE 4.11: Docking of tricin ligand with EGFR protein.

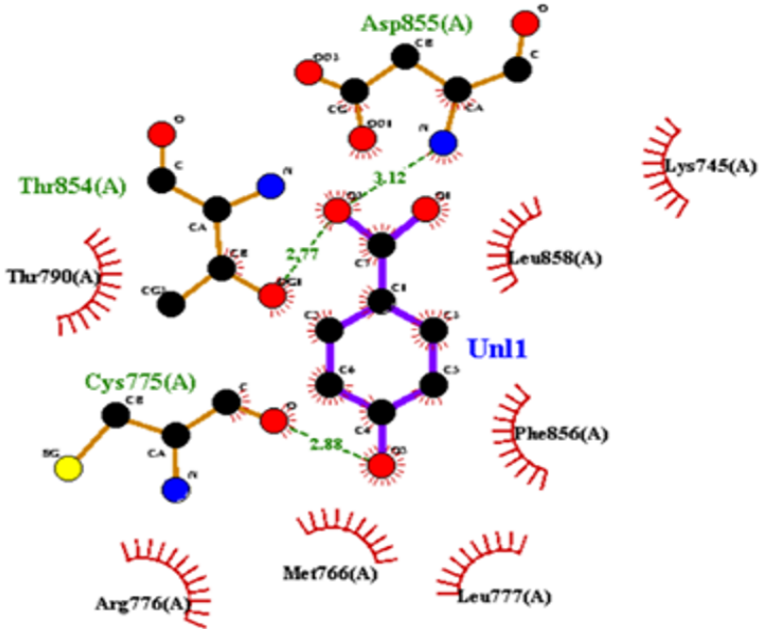


FIGURE 4.12: Docking of p-hydroxybenzoic ligand with EGFR protein.

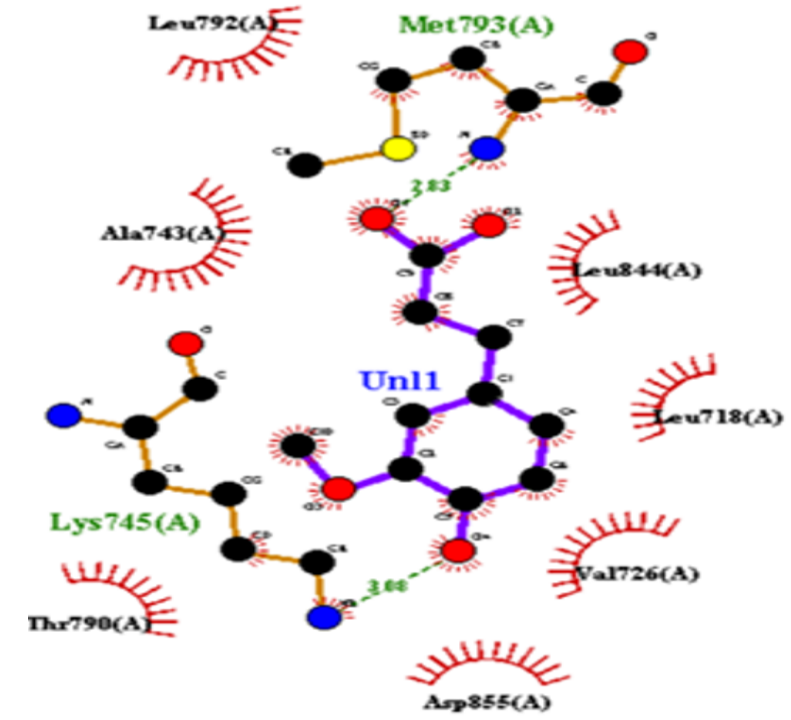


FIGURE 4.13: Docking of ferulic ligand with EGFR protein.

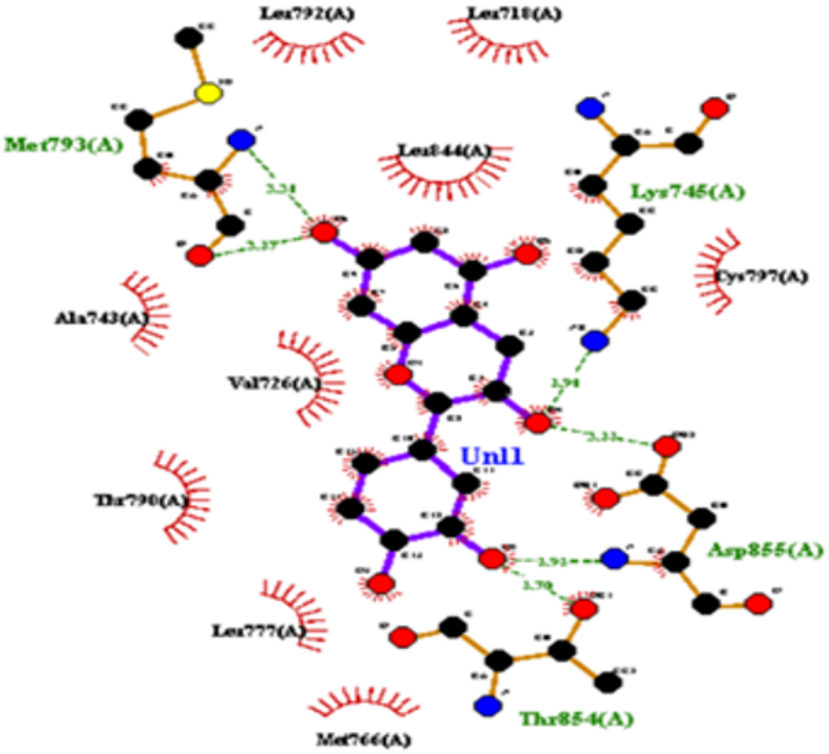


FIGURE 4.14: Docking of catechin ligand with EGFR protein.

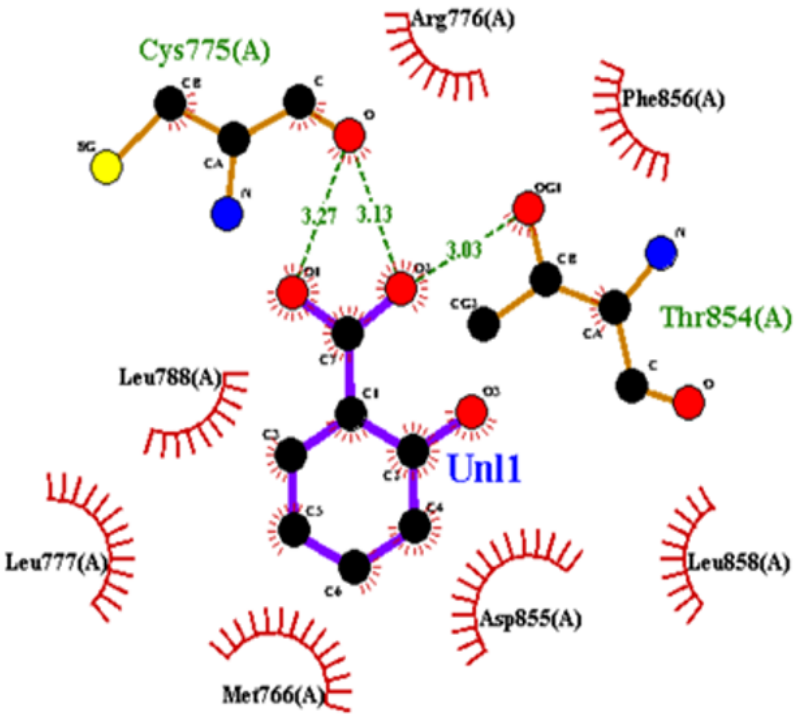


FIGURE 4.15: Docking of salicylic acids ligand with EGFR protein.

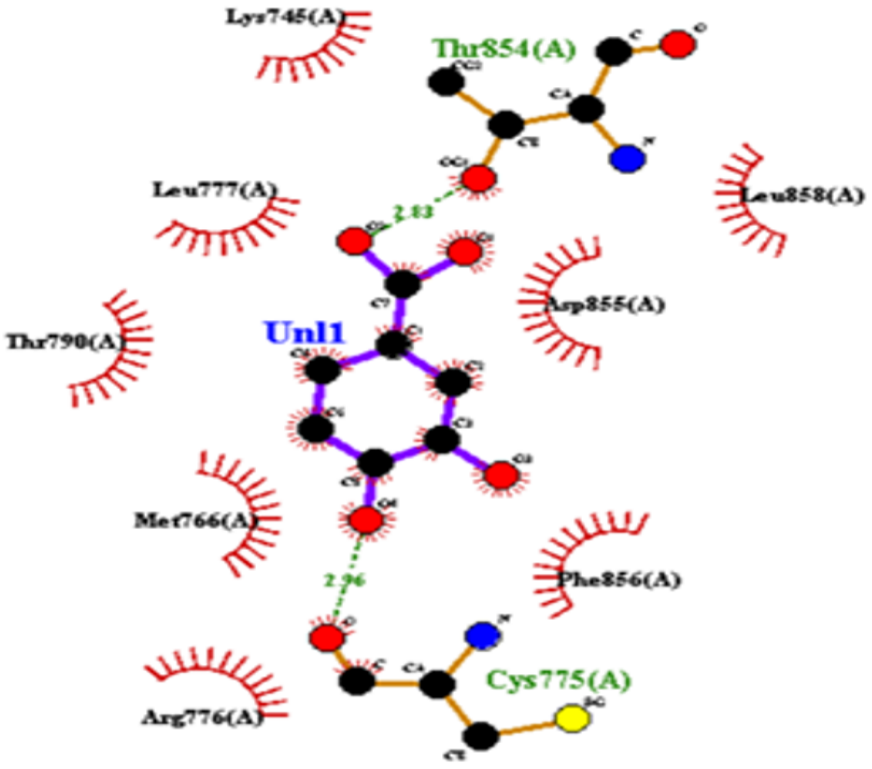


FIGURE 4.16: Docking of protocatechuic ligand with EGFR protein.

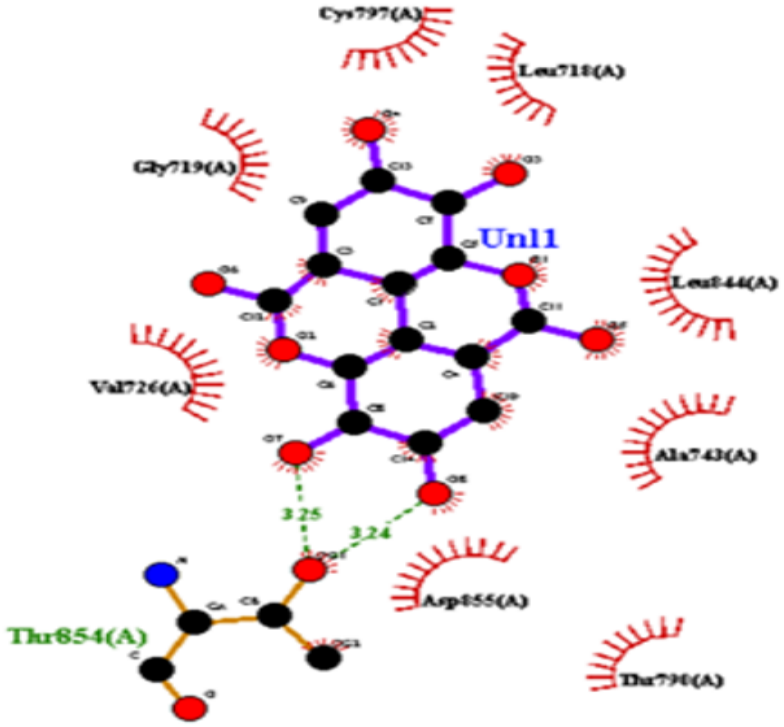


FIGURE 4.17: Docking of ellagic ligand with EGFR protein.

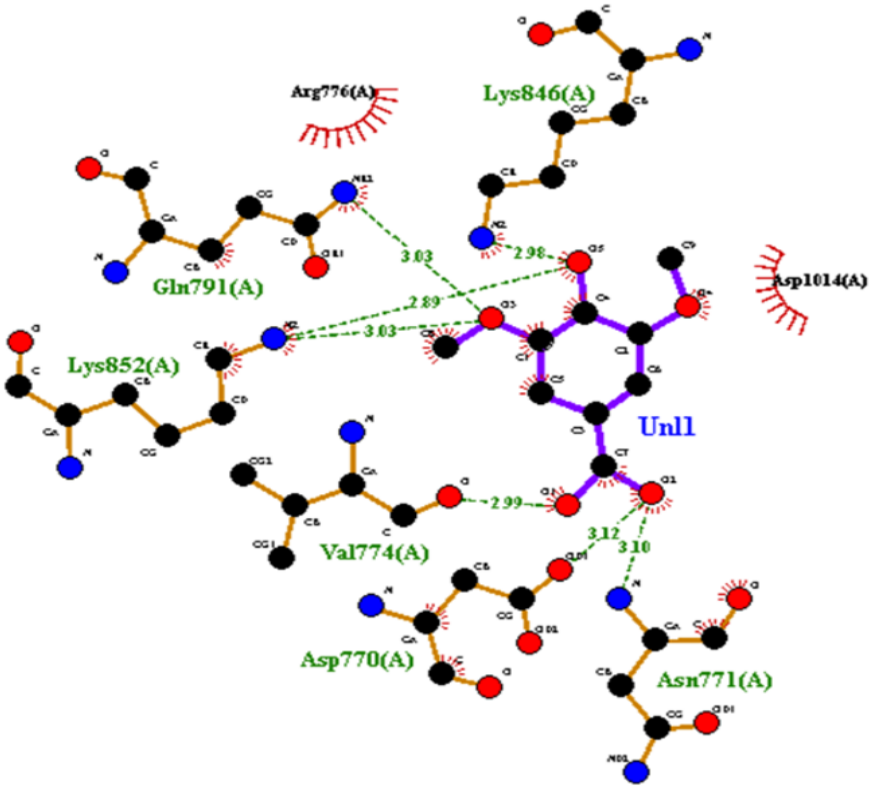


FIGURE 4.18: Docking of syringic ligand with EGFR protein.

Table 4.4: Interactions of ligands with EGFR protein.

Sr. No.	Ligand Names	Binding Energy	No. of HBs	Bonding Amino Acids	Distance	Hydro-Phobic Bonding
1	Acacetin	-8.1	2	O5-Met793-O, O5-Met793-N	2.69, 3.29	Leu844, Met1002, Ala743, Leu792, Gly796, Val726, Lys745, Leu858, Asp855, Thr854, Thr790, Met766, Leu788
2	Apigenin	-8.4	2	O5-Met793-N, O5-Met793-O	3.20, 2.83	Leu718, Leu792, Leu844, Thr790, Ala743, Thr854, Leu788, Asp855, Val726, Lys745, Met766
3	Caffeic acid	-6.5	3	O4-Cys775-O, O3-Phe856-O, O1-Leu788-O	3.01, 3.09, 3.08	Arg776, Met766, Thr854, Leu777, Asp855, Thr790, Lys745, Ala743, Val726, Ile744
4	Catechin	-8.3	6	O6-Met793-N, O6-Met793-O, O4-Lys745-NZ, O2-Asp855-N, O4-AspOD2, O2-Thr854-OG1	3.31, 3.27, 2.91, 2.92, 3.22, 2.70	Leu792, Leu718, Leu844, Cys797, Ala743, Val726, Thr790, Leu777, Met766

Table 4.4: (Continued).

Sr. No.	Ligand Names	Binding Energy	No. of HBs	Bonding Amino Acids	Distance	Hydro-Phobic Bonding
5	Chlorogenic acid	-8.0	7	O7-Asn842-OD1, O6-Asn842-OD1, O6-Arg841-O, O7-Asp855-OD2, O6-Asp855-OD2, O3-Cys775-C, O3-Met766-O	3.18, 2.70, 3.12, 2.84, 3.20, 3.04, 3.19	Leu844, Leu718, Val726, Ala743, Thr790, Lys745, Leu777, Thr854, Phe856, Arg776
6	Ellagic acid	-7.9	2	O7-Thr854-OG1, O8-Thr854-OG1	3.25, 3.24	Cys797, Leu718, Gly719, Leu844, Val726, Ala743, Asp855, Thr790
7	Ferulic acid	-6.1	2	O4-Lys745-NZ, O1-Met793-N	3.08, 2.83	Leu792, Ala743, Leu844, Leu718, Val726, Thr790, Asp855
8	Gallic acid	-6.3	4	O1-Thr854-OG1, O1-Asp855-N, O5-Met766-O, O4-Thr790-OG1	2.93, 3.26, 3.08, 2.71	Leu777, Cys775, Arg776, Phe856
9	Melilotic acid	-6.0	4	O3-Asp855-N, O1-Asp855-N, O3-Thr854-OG1, O1-Thr854-OG1	2.87, 3.30, 2.71, 2.83	Lys745, Val726, Leu777, Thr790, Arg776, Cys775
10	p-Hydroxy-benzoic acid	-5.9	3	O2-Asp855-N, O2-Thr854-OG1, O3-Cys775-O	3.12, 2.77, 2.88	Lys745, Thr790, Leu858, Phe856, Arg776, Met766, Leu777

Table 4.4: (Continued).

Sr. No.	Ligand Names	Binding Energy	No. of HBs	Bonding Amino Acids	Distance	Hydro-Phobic Bonding
11	Protocatechuic acid	-5.9	2	O2-Thr854-OG1, O4-Cys775-O	2.83, 2.96	Lys745, Leu777, Leu858, Asp855, Thr790, Met766, Phe856, Arg776
12	Salicylic acid	-6.0	3	O2-Cys775-O, O1-Cys775-O, O2-Thr854-OG1	3.13, 3.27, 3.03	Arg776, Phe856, Leu788, Leu777, Leu858, Asp855, Met766
13	Tricin	-7.8	4	O7-Lys745-N, O7-Ala743-O, O5-Thr854-OG1, O7-Leu788-O	3.02, 2.91, 3.08, 2.44	Thr790, Ile744, Ile789, Val726, Asp855, Arg841, Met793, Leu844, Gly796, Cys797
14	Vanillic acid	-5.5	4	O1-Thr854-OG1, O2-Asp855-OD1, O2-Asp855-N, O1-Asp855-N	2.70, 2.89, 3.30, 3.04	Ala743, Ile744, Thr790, Val726, Leu788, Lys745, Met766, Leu777, Leu858
15	Syringic acid	-5.3	7	O5-Lys846-NZ, O3-Gln791-NE2, O2-Asn771-N, O5-Lys852-NZ, O3-Lys852-NZ, O1-Val774-O, O2-Asp770-OD1	2.98, 3.03, 3.10, 2.89, 3.03, 2.99, 3.12	Arg776, Asp1014

4.5 Virtual Screening through Lipinski Rule of Five

It shows experimental and computational approach to evaluate permeability, soluble nature and effectiveness in process of discovery and development of drug [85]. The five rules are:

(1) H-bonds donors (OH, NH) < 5

(2) H-bonds acceptors (N, O etc) < 10

(3) LogP < 5 (LogP is the fat-water partition coefficient)

(4) Mol. Wt. < 500

(5) No. of rotatable keys < 10 [86].

According to results in table 4.5, acacetin, vanillic acid, syringic, salicylic, melilotic, tricin, catechin, apigenin, ferulic, gallic acid, p-hydroxyl benzoic acid, ellagic, caffeic and protocatechuic acid were obeying all rules expect chlorogenic acid which was violating one rule.

TABLE 4.5: Lipinski's Rule of Five applied to selected ligands.

Compounds	**Mol. Wt.** $< \mathbf{500}$	**Log P** $\leq \mathbf{5}$	**HBA** $< \mathbf{10}$	**HBD** $< \mathbf{5}$	**Surface Area**	**Rotatable Bonds**
Acacetin	284.267	2.8798	5	2	119.203	2
Syringic acid	198.174	1.1076	4	2	80.503	3
Salicylic acid	138.122	1.0904	2	2	57.546	1
Melilotic acid	166.176	1.4094	2	2	70.276	3
Tricin	330.292	2.5940	7	3	135.476	3
Catechin	290.271	1.5461	6	5	119.662	1
Chlorogenic acid	354.311	0.6459	8	6	141.587	4

Table 4.5: Lipinski's rule of five (Continued).

Compounds	**Mol. Wt.**	**Log P**	**HBA**	**HBD**	**Surface Area**	**Rotatable Bonds**
	< 500	**≤ 5**	**< 10**	**< 5**		
Apigenin	270.240	2.5768	5	3	112.519	1
Ferulic acid	194.186	1.4986	3	2	81.065	3
Ellagic acid	302.194	1.3128	8	4	118.565	0
Caffeic acid	180.159	1.1956	3	3	74.381	2
P-hydroxybenzoic acid	138.122	1.0904	2	2	57.546	1
Gallic acid	170.120	0.5016	4	4	67.135	1
Protocatechuic acid	154.121	0.7960	3	3	62.341	1
Vanillic acid	168.148	1.0990	3	2	69.025	2

4.6 ADMET Properties

In early stages of drug development, significant role is played by ADMET properties. This is because high quality drug candidates have both significant efficacy against target molecule as well as proper ADMET traits at required dose [87].

4.6.1 Absorption

The rate and extent to which the drug moves from its point of administration to point of action (target site) is designated as drug absorption. It is a very critical component in pharmacokinetics of drug. The drug must cross numerous barriers to reach its target site. The dose that should be administered is dictated by: 1) rate and extent of absorption, 2) time required by drug to produce a significant/ noticeable effect [88]. If the log Papp value of a chemical is greater than 0.90 cm/s,

then it is belived to have high-level Caco-2 permeability (on pkCSM) [89].

After oral drug administration, water solubility is critical part in pharmacological reaction of drug. Increase in water solubility ensures good drug features like high absorption and bioavailability. These in turn increase conc. of drug in plasma at target site, ensuring therapeutic effectiveness of drug [90]. A compound solubility ranges from -4 to -2; Soluble (-2 to 0), and extremely soluble (> 0). The skin permeability value (recommended > -2.5cm/h) is an important feature to enhance effectiveness of drug. It is specifically important in development of drugs with transdermal route of administration [89n]. P-glycoprotein (P-gp) is an ABC transporter (ATP-binding cassette). Its function is that of a biological barrier by elimination of toxins and xenobiotics. P-gp I/II inhibitor is the ability of chemical to stop transportation of P-gp I and P-gp II. There are important pharmacokinetics outcome for P-gp substrates when P-gp mediated transport is modified [89]. Salicylic acid and melilotic acid have high caco2 permebility. Syringic acid, tricin and catechin has low caco2 solubility and these are glycoprotein substrate (Table 4.6). Cholorogenic, ferulic, ellagic and caffeic has low caco2 solubility. Whereas chlorogenic, apigenin and ellagic are glycoprotein substrate (Table 4.7). P-hydroxybenzoic acid and acacetin have high caco2 permeability. Protechuic, gallic and vanillic has low caco2 permeability. Acacetin is glycoprotein substrate (Table 4.8). According to the absorption properties of ligands (Table 4.6 – Table 4.8), all had low solubility in water, low permeability in skin. The melilotic acid had highest Caco-2 permeability and acacetin exhibited highest intestinal absorption.

TABLE 4.6: Ligands (syringic acid, salicylic acid, melilotic acid, tricin and catechin) with their respective absorption properties. (Yes= + ve, No = - ve).

Model Name	**Syringic acid**	**Salicylic acid**	**Melilotic acid**	**Tricin**	**Catechin**
Water solubility	-2.223	-1.808	-2.419	-3.276	-3.117
Permeability of Caco-2	0.495	1.151	1.203	0.120	-0.283
Absorption in human intestine (%)	73.076	83.887	93.682	89.713	68.829

Table 4.6: (Continued).

Model Name	**Syringic acid**	**Salicylic acid**	**Melilotic acid**	**Tricin**	**Catechin**
Permeability of skin	-2.735	-2.723	-2.711	-2.735	-2.735
P - glycoprotein substrate	+ ve	- ve	- ve	+ ve	+ ve
P - glycoprotein I inhibitor	- ve	- ve	- ve	- ve	- ve
P - glycoprotein II inhibitor	- ve	- ve	- ve	- ve	- ve

TABLE 4.7: Ligands (chlorogenic, apigenin, ferulic, ellagic, caffeic acid) with their respective predicted absorption properties. (Yes= + ve, No = - ve).

Model Name	**Chlorogenic acid**	**Apigenin**	**Ferulic acid**	**Ellagic acid**	**Caffeic acid**
Water solubility	-2.449	-3.329	-2.817	-3.181	-2.33
Permeability of Caco-2	-0.84	1.007	0.176	0.335	0.634
Absorption in human intestine (%)	36.377	93.25	93.685	86.684	69.407
Permeability of skin	-2.735	-2.735	-2.72	-2.735	-2.722
P - glycoprotein substrate	+ ve	+ ve	- ve	+ ve	- ve
P - glycoprotein I inhibitor	- ve	- ve	- ve	- ve	- ve
P - glycoprotein II inhibitor	- ve	- ve	- ve	- ve	- ve

TABLE 4.8: Ligands (P-hydroxybenzoic, protocatechuic, acacetin, gallic and vanillic acid) with their respective absorption traits. (Yes= + ve, No = - ve).

Model Name	**P-hydroxybenzoic acid**	**Protocatechuic acid**	**Acacetin**	**Gallic acid**	**Vanillic acid**
Water solubility	-1.877	-2.069	-3.284	-2.56	-1.838

Table 4.8: (Continued).

Model Name	**P-hydroxybenzoic acid**	**Protocatechuic acid**	**Luteolin**	**Gallic acid**	**Vanillic acid**
Permeability of Caco-2	1.151	0.49	1.137	-0.081	0.33
Absorption in human intestine (%)	83.961	71.174	94.318	43.374	78.152
Permeability of skin	-2.723	-2.727	-2.737	-2.735	-2.726
P-glycoprotein substrate	- ve	- ve	+ ve	- ve	- ve
P-glycoprotein I inhibitor	- ve	- ve	- ve	- ve	- ve
P-glycoprotein II inhibitor	- ve	- ve	- ve	- ve	- ve

4.6.2 Distribution

The drug is distributed to interstitial and intracellular compartments when the drug reached the blood stream. Many drugs may bound to plasma protein in blood. This binding is reversible and the bound drug and unbound drug exist in dynamic equilibrium. Any change in conc. of one of them is always followed by change in conc. of other. Only unbound drug is considered pharmacologically active and have the ability to cross membranes and/ or interact with target sites.

The volume of distribution (Vd) represents the vol. in which a drug needs to distribute to achieve the same conc. as that detected in the blood plasma. We can calculate Vd by division of amount of the drug in the body with the plasma conc. Vd gives an insight about how much drug is present in extravascular tissues. Vd is affected by lipid solubility, since more lipid soluble drugs have better cell penetration (and more Vd) [88].

The drug capacity to be dispersed in the body is determined by 2 important factors i.e. volume of supply at steady state (VDss) and the blood-brainbarrier (BBB). When a molecule has VDss higher than 0.45, it has good dispersion. When log BB is higher than 0.3, molecule can move quickly across BBB [89].

All these ligands show fraction unbound and these ligands cannot cross BBB and CNS. According to the Table 4.9 - 4.11), all the ligands show fraction unbound and these ligands cannot cross BBB and CNS. All these compounds have low VDss except apigenin, tricin and catechin.

TABLE 4.9: Distribution properties predicted of ligands (syringic acid, salicylic acid, melilotic acid, tricin and catechin).

Model Name	**Syringic acid**	**Salicylic acid**	**Melilotic acid**	**Tricin**	**Catechin**
Vd in human	-1.443	-1.57	-1.166	0.798	1.027
Unbound fraction (human)	0.601	0.563	0.432	0.084	0.235
Permeability of BBB	-0.191	-0.334	-0.213	-1.115	-1.054
Permeability of CNS	-2.701	-3.21	-2.535	-3.411	-3.298

TABLE 4.10: Distribution properties predicted of ligands (Chlorogenic acid, apigenin, ferulic acid, ellagic and caffeic).

Model Name	**Chlorogenic acid**	**Apigenin**	**Ferulic acid**	**Ellagic acid**	**Caffeic acid**
Vd in human	0.581	0.822	-1.367	0.375	-1.098
Unbound fraction (human)	0.658	0.147	0.343	0.083	0.529
Permeability of BBB	-1.407	-0.734	-0.239	-1.272	-0.647
Permeability of CNS	-3.856	-2.061	-2.612	-3.533	-2.608

TABLE 4.11: Distribution properties predicted of ligands (p-hydroxybenzoic, protocatechuic acid, acacetin acid, gallic acid and vanillic acid).

Model Name	**P-hydroxy-benzoic acid**	**Protoca-techuic acid**	**Acacetin**	**Gallic acid**	**Vanillic acid**
Vd in human	-1.557	-1.298	0.346	-1.855	-1.739
Unbound fraction (human)	0.592	0.648	0.080	0.617	0.518
Permeability of BBB	-0.334	-0.683	-0.196	-1.102	-0.380
Permeability of CNS	-3.210	-3.305	-2.159	-3.740	-2.628

4.6.3 Metabolism

After distribution of drug throughout the body, it is metabolized. It converts to polar inactive metabolites so that it can be eliminated from the body. Drugs which are lipophilic in nature can pass through biological membranes and reach target site with ease. This lipophilicity can hinder their excretion from the body [88]. An enzyme that is involved in detoxification in liver is cytochrome P450. It also plays role in drug metabolism. The pharmacokinetics of drugs is significantly impacted by P450 inhibitors. Therefore, it is important to evaluate if the molecule in question is substrate of CYP2D6/CYP3A4, and can be acted upon by P450. CYP 450 acts on xenobiotics, oxidizing them and allowing their excretion [89]. According to table 4.12, only tricin have CYP1A2 and CYP2C19 inhibitors, rest all ligands have no substrate and as well as no inhibitors. According to table 4.13, chlorogenic acid, apigenin, ferulic acid, ellagic and caffeic ligands have neither CYP2D6, CYP3A4 substrates nor they have inhibitors. Apigenin has CYP1A2 and CYP2C19 inhibitors. Ellagic acid has CYP1A2 inhibitor. According to table 4.14, P-hydroxybenzoic, protocatechuic, gallic, vanillic are neither have substrate nor inhibitors. Acacetin has CYP3A4 substrate and CYP1A2, CYP2C19, CYP2C9, CYP3A4 inhibitors.

According to the tables (4.12 – 4.14), only acacetin is a substrate of CYP3A4. Most of the ligands are neither substrate nor inhibitor. Some are inhibitors as well.

TABLE 4.12: Metabolism of ligands (syringic acid, salicylic acid, melilotic acid, tricin and catechin). ((+) = Yes, (-) = No).

Model Name	**Syringic acid**	**Salicylic acid**	**Melilotic acid**	**Tricin**	**Catechin**
Substrate of CYP2D6	(-)	(-)	(-)	(-)	(-)
Substrate of CYP3A4	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP1A2	(-)	(-)	(-)	(+)	(-)
Inhibitor of CYP2C19	(-)	(-)	(-)	(+)	(-)
Inhibitor of CYP2C9	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP2D6	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP3A4	(-)	(-)	(-)	(-)	(-)

TABLE 4.13: Metabolism of ligands (Chlorogenic acid, apigenin, ferulic acid, ellagic and caffeic). ((+) = Yes, (-) = No).

Model Name	**Chlorogenic acid**	**Apigenin**	**Ferulic acid**	**Ellagic acid**	**Caffeic acid**
Substrate of CYP2D6	(-)	(-)	(-)	(-)	(-)

Table 4.13: (Continued).

Model Name	**Chlorogenic acid**	**Apigenin**	**Ferulic acid**	**Ellagic acid**	**Caffeic acid**
Substrate of CYP3A4	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP1A2	(-)	(+)	(-)	(+)	(-)
Inhibitor of CYP2C19	(-)	(+)	(-)	(-)	(-)
Inhibitor of CYP2C9	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP2D6	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP3A4	(-)	(-)	(-)	(-)	(-)

TABLE 4.14: Metabolism of ligands (p-hydroxybenzoic, protocatechuic acid, acacetin acid, gallic acid and vanillic acid). ((+) = Yes, (-) = No).

Model Name	**P-hydroxy-benzoic acid**	**Protocate-chuic acid**	**Acacetin**	**Gallic acid**	**Vanillic acid**
Substrate of CYP2D6	(-)	(-)	(-)	(-)	(-)
Substrate of CYP3A4	(-)	(-)	(+)	(-)	(-)
Inhibitor of CYP1A2	(-)	(-)	(+)	(-)	(-)
Inhibitor of CYP2C19	(-)	(-)	(+)	(-)	(-)

Table 4.14: (Continued).

Model Name	**P-hydroxy-benzoic acid**	**Protocat-echuic acid**	**Acacetin**	**Gallic acid**	**Vanillic acid**
Inhibitor of CYP2C9	(-)	(-)	(+)	(-)	(-)
Inhibitor of CYP2D6	(-)	(-)	(-)	(-)	(-)
Inhibitor of CYP3A4	(-)	(-)	(+)	(-)	(-)

4.6.4 Excretion

Liver metabolism and renal excretion are two main way via which drugs are excreted from the body. Some drugs may also be secreted in negligible amount via sweat, tears and saliva. The elimination of drug via these routes depends on 1) pH od urine and 2) nonionized lipophilic diffusion of drug through glandular epithelial cells [88]. Total clearance of above mentioned ligands ranges from 0.183 to 0.663 (Table 4.15 - 4.17). All these ligands have no renal OCT2 substrate and total clearance is given accordingly.

TABLE 4.15: Excretion of ligands (syringic acid, salicylic acid, melilotic acid, tricin and catechin).

Model Name	**Syringic acid**	**Salicylic acid**	**Melilotic acid**	**Tricin**	**Catechin**
Total excretion	0.646	0.607	0.315	0.62	0.183
Substrate of Renal OCT2	Nil	Nil	Nil	Nil	Nil

TABLE 4.16: Excretion of ligands (Chlorogenic acid, apigenin, ferulic acid, ellagic and caffeic).

Model Name	**Chlorogenic acid**	**Apigenin**	**Ferulic acid**	**Ellagic acid**	**Caffeic acid**
Total excretion	0.307	0.566	0.623	0.537	0.508
Substrate of Renal OCT2	Nil	Nil	Nil	Nil	Nil

TABLE 4.17: Excretion of ligands (p-hydroxybenzoic, protocatechuic acid, acacetin acid, gallic acid and vanillic acid)..

Model Name	**P-hydroxy-benzoic acid**	**Protocat-echuic acid**	**Acacetin**	**Gallic acid**	**Vanillic acid**
Total excretion	0.593	0.551	0.663	0.518	0.628
Substrate of Renal OCT2	Nil	Nil	Nil	Nil	Nil

4.6.5 Toxicity Prediction

LD50 is the toxic dose decided in mg/kg body weight. LD refers to the median lethal dose, indicating the amount of a compound that causes death in 50% of test subjects. *T. pyroformis* toxicity values greater than -0.5 are considered non-toxic [91].

Using *Salmonella typhimurium,* AMES test is employed to analyze mutagenicity of a compound. The bacterial cells upon exposure to mutagenic materials undergo several mutations, and allow bacteria to prosper in environment with no histidine. Positive results implicate that compound is mutagenic and may cause cancer. hERG I and II sheds light on cardiotoxicity as they lead to lethal arrythmias of ventricles by inhibiting potassium channels. Pharmacokinetics of drug is affected

by hepatotoxicity. It holds significant importance in drug development and is one of leading reasons for drug attrition. The oral rat chronic toxicity (dose 0.341 to 2.674) analyzes the impact of drug on body upon repeated ingestion, skin application or inhalation during a specific time period. This parameter allows evaluation of MRTD to start clinical trial on humans. MRTD estimates a threshold where substance becomes toxic to humans and is crucial in dose determination during clinical trials [92]. Minnow toxicology testing determines the amount/dose which causes 50% death in research animals (fathead minnows). Substances have acute toxicity when they have low LC50 (less than 0.5 mM or Log LC50 less than -0.3). Skin sensitization determines if the substance will cause allergic contact dermatitis [93].

Melilotic and ferulic acid shows no AMES toxicity and both has similar maximal tolerated dose and no hERG I and hERG II inhibition. Both compounds shows no hepatoxicity and skin sensitization (Table 4.18).

Both caffeic acid and hydroxy benzoic acid shows no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.19).

Both syringic acid and acacetin show no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.20).

Both tricin and catechin show no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.21). Chlorogenic acid and salicylic acid shows no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.22).

Apigenin and ellagic acid show no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.23). Gallic acid, vanillic acid and protocatechuic acid shows no AMES toxicity and no hERG I and hERG II inhibition. Both compounds show no hepatoxicity and skin sensitization (Table 4.24).

TABLE 4.18: Toxicity prediction of melilotic acid and ferulic acid. (Neg = Negative).

Model Name	**Melilotic Acid**	**Ferulic Acid**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	1.092	1.082
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.168	2.282
Oral Chronic Toxicity in Rat (LOAEL)	2.555	2.065
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.319	0.271
Toxicity in minnow	1.556	1.825

TABLE 4.19: Toxicity prediction of caffeic acid and hydroxy benzoic acid. (Neg = Negative).

Model Name	**Caffeic Acid**	**P-hydroxy Benzoic Acid**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	1.145	0.846
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.383	2.255
Oral Chronic Toxicity in Rat (LOAEL)	2.092	2.483
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.293	0.268
Toxicity in minnow	2.246	1.812

TABLE 4.20: Toxicity prediction of syringic acid and acacetin acid. (Neg = Negative).

Model Name	**Syringic Acid**	**Acacetin**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	1.374	0.09
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.157	2.22
Oral Chronic Toxicity in Rat (LOAEL)	2.415	1.259
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.281	0.422
Toxicity in minnow	2.554	1

TABLE 4.21: Toxicity prediction of tricin and catechin. (Neg = Negative).

Model Name	**Tricin**	**Catechin**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	0.351	0.438
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.229	2.428
Oral Chronic Toxicity in Rat (LOAEL)	1.82	2.5
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.329	0.347
Toxicity in minnow	1.754	3.358

TABLE 4.22: Toxicity prediction of chlorogenic acid and salicylic acid. (Neg = Negative).

Model Name	**Chlorogenic acid**	**Salicylic acid**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	-0.134	0.61
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	1.973	2.282
Oral Chronic Toxicity in Rat (LOAEL)	2.982	2.483
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.285	0.263
Toxicity in minnow	5.741	1.812

TABLE 4.23: Toxicity prediction of apigenin and ellagic acid. (Neg = Negative).

Model Name	**Apigenin**	**Ellagic Acid**
Toxicity of AMES	Neg	Neg
Highest Tolerable Dose (human)	0.328	0.476
Inhibitor of hERG I	Neg	Neg
Inhibitor of hERG II	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.45	2.399
Oral Chronic Toxicity in Rat (LOAEL)	2.298	2.698
Liver toxicity	Neg	Neg
Dermal sensitivity	Neg	Neg
Toxicity in *T. pyriformis*	0.38	0.295
Toxicity in minnow	2.432	2.11

TABLE 4.24: Toxicity prediction of gallic acid, protocatechuic acid and vanillic acid. (Neg = Negative).

Model Name	**Gallic Acid**	**Protocatechuic Acid**	**Vanillic Acid**
Toxicity of AMES	Neg	Neg	Neg
Highest Tolerable Dose (human)	0.7	0.814	0.179
Inhibitor of hERG I	Neg	Neg	Neg
Inhibitor of hERG II	Neg	Neg	Neg
Oral Acute Toxicity in Rat (LD50)	2.218	2.423	2.454
Oral Chronic Toxicity in Rat (LOAEL)	3.06	2.021	2.032
Liver toxicity	Neg	Neg	Neg
Dermal sensitivity	Neg	Neg	Neg
Toxicity in *T. pyriformis*	0.285	0.273	0.265
Toxicity in minnow	3.188	2.451	1.926

4.7 Lead Compound Identification

Lipinski's Ro5 and ADMET properties and toxicity prediction as primary filter while binding scores as secondary filter was applied to identify the lead compound. Chlorogenic acid was knocked out for violating one rules of Lipinski while rest of the ligands were following the rules.

In secondary filter screening ADMET properties were observed. Syringic acid, tricin, ferulic, ellagic, caffeic, catechin, Protocatechuic, gallic and vanillic were knocked out for low CaCO2 permeability. Furthermore, Apigenin was knocked out as it is the inhibitor of CYP1A2, CYP2C19.

After primary and secondary filter screening remaining compound acacetin was chosen as the best suitable lead compound as its docking score is -8.1. Second highest after apigenin. It is also a substrate of Cytochrome P450 enzyme, and also have 13 interactions (hydrophobic in nature) with the protein which will makes the interaction even stronger.

4.8 Drug Selection

The standard drug was selected from KEGG disease database.

4.8.1 Gefitinib

One of the primary treatment of NSCLC is gefitinib; an inhibitor of EGFR. It specifically acts on ATP binding region of tyrosine kinase present in EGFR. This prevents EGFR from getting phosphorylated and the downstream signals are blocked. The outcome is inhibition of cell growth, specialization and survival [38].

4.8.2 Effects on Body

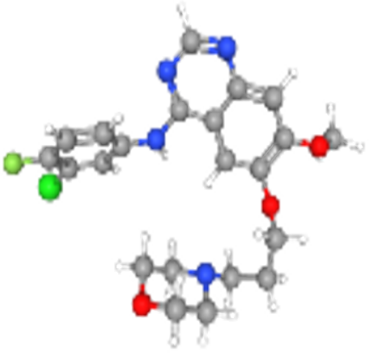
The tolerance of gefitinib in phase I clinical trials was quite well. It showed efficacy even below the maximum tolerated dose. The side effects are mild to moderate, can be managed with proper intervention and are reversible. These include dry skin, diarrhea, vomiting, nausea and acneiform rash.

There were 1% cases that reported of interstitial lung disease (like inflammation/ pneumonia). Other more serious side effects included swelling of the lips and tongue, difficulty of breathing, allergic reactions, elevated liver enzymes [94]. Chemical and structural information of gefitinib can be noted in Table 4.25.

TABLE 4.25: Toxicity prediction of gallic acid, protocatechuic acid and vanillic acid. (Neg = Negative).

Property	**Value**
PubChem Id	123631
Molecular Weight	446.9 g/mol
Rotatable Bonds	8
Formula	C22H24ClFN4O3

Table 4.25: (Continued).

Property	**Value**
Structure	

4.9 Physiochemical Properties

Physicochemical properties describe the physical and chemical properties of a molecule (protein, ligand, drug etc.). These properties of gefitinib can be noted from Table 4.26.

TABLE 4.26: Chemical and physical properties of gefitinib.

Descriptor	**Value**
Mol. Wt.	446.91
LogP	4.2756
Rotatable Bonds	8
Acceptors	7
Donors	1
Surface Area	184.642

4.10 Drug Docking

Gefitinib was docked with EGFR protein. The docking score can be noted in Table 4.27.

TABLE 4.27: Docking of gefitinib with EGFR protein.

Compound	Binding Score	Cavity Size	Grid Map	HBD	HBA	Mol. Wt.	Log P	Rotatable Bonds
Gefitinib	-8.9	2263	37	1	7	446.91	4.2756	8

4.11 Drug Protein Interaction

Ligplot (v1.4.5) was used to determine the interactions of gefitinib (drug) with EGFR (target protein). These interactions can be noted in Table 4.28 and Figure 4.19.

TABLE 4.28: Gefitinib interaction with EGFR target protein.

Sr No	Ligand Names	Binding Energy	No of HBs	Bonding Amino Acids	Distance (Å)	Hydro-Phobic Bonding
1	Gefitinib	-8.9	4	N2-Arg836-NH2, O3-His893-NE2, N4-Tyr891-O, O2-Lys960-NZ	3.12, 3.16, 2.89, 3.00	0

4.12 Gefitinib Toxicity Prediction

The toxicity prediction of gefitinib can be noted in Table 4.29.

TABLE 4.29: Toxicity prediction of gefitinib.

Sr No	Model Name	Predicted Values
1	Toxicity of AMES	No
2	Highest Tolerable Dose (human)	0.011

Table 4.29: (Continued).

Sr No	Model Name	Predicted Values
3	Inhibitor of hERG I	No
4	Inhibitor of hERG II	Yes
5	Oral Acute Toxicity in Rat (LD50)	2.859
6	Oral Chronic Toxicity in Rat (LOAEL)	1.334
7	Liver toxicity	Yes
8	Dermal sensitivity	No
9	Toxicity in *T. pyriformis*	0.303
10	Toxicity in minnow	-0.868

4.13 Drug ADME Properties

Absorption of gefitinib can be noted in Table 4.30.

Table 4.30: Absorption prediction of gefitinib.

Sr. No.	Parameters	Predicted Value
1	Solubility in water	-4.124
2	Permeability in Caco2	1.246
3	Absorption in human intestine	90.992
4	Permeability in skin	-2.749
5	P - glycoprotein substrate	Yes
6	P - glycoprotein I inhibitor	Yes
7	P - glycoprotein II inhibitor	Yes

4.14 Distribution

Distribution properties of gefitinib were predicted and can be noted in Table 4.31.

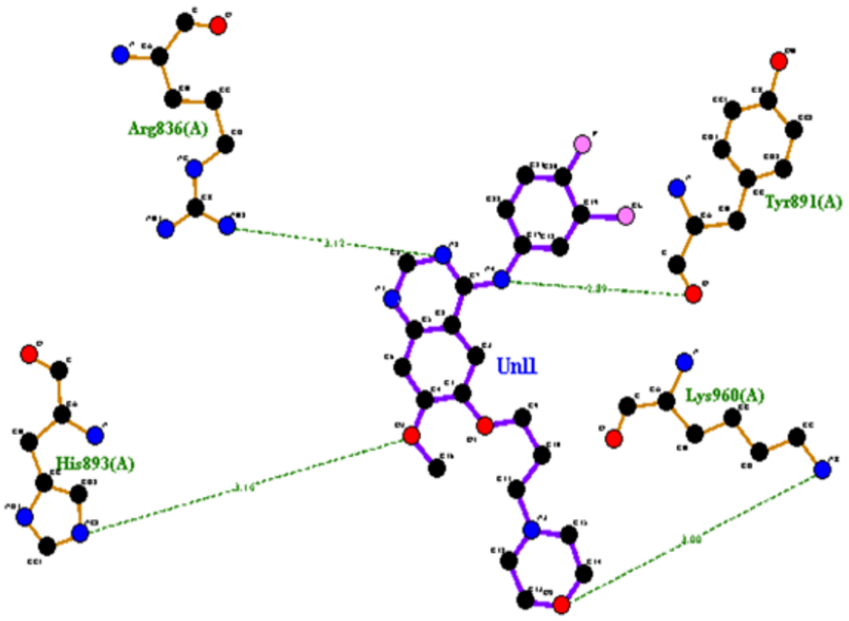


FIGURE 4.19: Interactions developed between gefitinib drug and EGFR target protein.

TABLE 4.31: Distribution prediction of gefitinib.

Sr. No.	Model Name	Predicted Value
1	VDss (human)	1.08
2	Unbound fraction (human)	0.111
3	Permeability to BBB	-0.738
4	Permeability to CNS	-3.198

4.14.1 Metabolism

Metabolism properties of gefitinib were predicted and can be noted in Table 4.32.

TABLE 4.32: Metabolism properties prediction of gefitinib drug.

Sr. No.	Model Name	Predicted Value
1	Substrate of CYP2D6	Negative
2	Substrate of CYP3A4	Positive

Table 4.32: (Continued).

Sr. No.	Model Name	Predicted Value
3	Inhibitor of CYP1A2	Positive
4	Inhibitor of CYP2C19	Negative
5	Inhibitor of CYP2C9	Negative
6	Inhibitor of CYP2D6	Negative
7	Inhibitor of CYP3A4	Positive

4.14.2 Excretion

Metabolism properties of gefitinib were predicted and can be noted in Table 4.33.

TABLE 4.33: Excretion properties prediction of gefitinib drug.

Sr. No.	Model Name	Predicted Value
1	Total Clearance	0.937
2	Renal OCT2 Substrate	Yes

4.15 Comparison with Lead Compound

Gefitinib and acacetin exhibit distinct physiochemical and ADMET properties. Gefitinib, with a higher molecular weight and LogP, suggests greater lipophilicity and potentially higher bioavailability. However, acacetin demonstrates better water solubility and comparable intestinal absorption. Both compounds exhibit P-glycoprotein substrate and inhibitor properties, though gefitinib shows broader inhibition. Distributionally, gefitinib has a higher volume of distribution. Metabolically, both are CYP3A4 substrates, but gefitinib is a CYP1A2 inhibitor while acacetin inhibits CYP1A2, CYP2C19, and CYP2C9. Excretionally, gefitinib has a slightly higher total clearance. Toxicity-wise, gefitinib shows higher hepatotoxicity and hERG II inhibition, while acacetin displays no toxicity in most in vitro and in vivo models.

Acacetin, despite lower potency, exhibits a more favorable drug-like profile compared to gefitinib with better physicochemical properties and a potentially no toxicity profile. Gefitinib shows no hydrophobic interactions while acacetin shows 13 hydrophobic interactions.

4.15.1 Physiochemical Properties Comparison

Comparison of physicochemical properties of gefitinib and acacetin in table 4.34.

TABLE 4.34: Physicochemical properties comparison of gefitinib and acacetin.

Descriptor	**Gefitinib**	**Acacetin**
Mol. Wt.	446.91	284.267
LogP	4.2756	2.8798
Rotatable Bonds	8	2
Acceptors	7	5
Donors	1	2
Surface Area	184.642	119.203

4.15.2 ADMET Comparision

Gefitinib and acacetin ADMET properties were compared. These included absorption (Table 4.35), distribution (Table 4.36), metabolism (Table 4.37), excretion (Table 4.38) and toxicity (Table 4.39).

TABLE 4.35: Absorption properties comparison of gefitinib and acacetin.

Sr. No.	**Model Name**	**Gefitinib**	**Acacetin**
1.	Water solubility	-4.124	-3.284
2.	Caco2 permeability	1.246	1.137
3.	Intestinal absorption (human)	90.992	94.318
4.	Permeability of skin	-2.749	-2.737

Table 4.35: (Continued).

Sr No.	Model Name	Gefitinib	Acacetin
5.	P - glycoprotein substrate	Affirmative	Affirmative
6.	P - glycoprotein I inhibitor	Affirmative	Negative
7.	P - glycoprotein II inhibitor	Affirmative	Negative

Table 4.36: Distribution properties comparison of gefitinib and acacetin.

Sr no.	Model Name	Gefitinib	Acacetin
1.	VDss (human)	1.08	0.346
2.	Unbound fraction in human	0.111	0.08
3.	Permeable to BBB	-0.738	-0.196
4.	Permeable to CNS	-3.198	-2.195

Table 4.37: Metabolism properties comparison of gefitinib and acacetin. (✓= Yes, × = No).

Sr no.	Model Name	Gefitinib	Acacetin
1.	Substrate of CYP2D6	×	×
2.	Substrate of CYP3A4	✓	✓
3.	Inhibitor of CYP1A2	✓	✓
4.	Inhibitor of CYP2C19	×	✓
5.	Inhibitor of CYP2C9	×	✓
6.	Inhibitor of CYP2D6	×	×
7.	Inhibitor of CYP3A4	✓	✓

Table 4.38: Excretion properties comparison of gefitinib and acacetin.

Sr No.	Model Name	Gefitinib	Acacetin
1.	Total excretion	0.937	0.663
2.	Substrate of renal OCT2	Yes	No

TABLE 4.39: Toxicity properties comparison of gefitinib and acacetin.

Sr No.	**Model Name**	**Gefitinib**	**Acacetin**
1.	Toxicity of AMES	Negative	Negative
2.	Highest Tolerable Dose (human)	0.011	0.09
3.	Inhibitor of hERG I	Negative	Negative
4.	Inhibitor of hERG II	Positive	Negative
5.	Oral Acute Toxicity in Rat (LD50)	2.859	2.22
6.	Oral Chronic Toxicity in Rat (LOAEL)	1.334	1.259
7.	Liver toxicity	Positive	Negative
8.	Dermal sensitivity	Negative	Negative
9.	Toxicity in *T. pyriformis*	0.303	0.422
10.	Toxicity in minnow	-0.868	1

4.15.3 Comparison of Docking Interactions

The comparison of docking interactions of gefitinib and acacetin with target protein EGFR (Table 4.40).

TABLE 4.40: Comparing docking interactions developed by gefitinib and acacetin with target protein EGFR.

Sr No	**Ligand Names**	**Binding Energy**	**No of HBs**	**Bonding Amino Acids**	**Distance (Å)**	**Hydro-Phobic Bonding**
1	Gefitinib	-8.9	4	N2-Arg836-NH2, O3-His893-NE2, N4-Tyr891-O, O2-Lys960-NZ	3.12, 3.16, 2.89, 3.00	0

Table 4.40: (Continued).

Sr No	**Ligand Names**	**Binding Energy**	**No of HBs**	**Bonding Amino Acids**	**Distance (Å)**	**Hydro-Phobic Bonding**
2	Acacetin	-8.1	2	O5-Met793-O, O5-Met793-N	2.69, 3.29	Leu844, Met1002, Ala743, Leu792, Gly796, Val726, Lys745, Leu858, Asp855, Thr854, Thr790, Met766, Leu788

Various studies highlight the potential therapeutic role of acacetin in treatment of various kinds of cancers like prostate, cervix, liver, blood and mouth. Studies reported its mechanism of action in which major role is played by methoxy group at C-4' on the B-ring and the 5,7-dihydroxy moieties on the A-ring of the molecule. These inhibit ethoxyresorufin-O-deethylase.

Acacetin showed promise for chemoprevention by inhibition of ATP-competitive PI3-K. An in vivo study reported that the compound has the potential to decrease growth of SK-MEL-28 tumours. It also has the ability to prevent Akt phosphorylation (group of enzymes directly involved in signal transmission, cell growth and cell survival).

Cell transformation induced by EGF was suppressed by acacetin, reducing Akt and p70S6K phosphorylation. This lead to inhibition of PI3-K activity. Furthermore,

it blocked anchorage-dependent as well as anchorage-independent growth of cells expressing PI3-K activity, and blocked phosphorylation of tyrosine of Stat-1 and -3 (these induce signals and activate transcription). Studies conducted in vitro, ex vivo and in vivo have exhibited its ability to block Stat-signaling and effectively suppressing angiogenesis. This highlights the potential therapeutic ability of acacetin as tumour growth inhibitor [95].

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

The objective of this work was to identify phytocompounds as inhibitor of EGFR in NSCLC. Fifteen compounds were selected for docking against EGFR protein. Virtual screening of compounds was done using Lipinski rule of 5 and ADMET analysis through PkCsm tool. After that docking was done using CB dock and visualization of docking results were done using LIGPLOT plus.

Acacetin was selected as lead compound because it follows Lipinski's RO5 and it has favorable ADMET properties and physiochemical properties and has more good hydrophobic interactions as compared to gefitinib. Since virtual drug design went well, more in vitro and in vivo research is suggested to confirm acacetin's potential as EGFR inhibitor in NSCLC.

5.2 Recommendations

1. Further experimental validation of acacetin's EGFR inhibitory activity through in vitro and in vivo studies is necessary to confirm its therapeutic potential in

NSCLC.

2. Optimization of acacetin for enhanced bioavailability and specificity towards EGFR could improve its efficacy in treating NSCLC.

3.Exploration of combination therapies involving acacetin and existing EGFR inhibitors may provide synergistic effects for better management of NSCLC.

Bibliography

[1] K. Xu et al., "Progress of exosomes in the diagnosis and treatment of lung cancer," *Biomedicine Pharmacotherapy*, vol. 134, p. 111111, Dec. 2020.

[2] K. Wadowska, I. Bil-Lula, Trembecki, and M. Śliwińska Mossoń, "Genetic markers in lung cancer diagnosis: a review," *International Journal of Molecular Sciences*, vol. 21, p. 4569, Jun. 2020.

[3] H. Lemjabbar-Alaoui, O. U. Hassan, Y.-W. Yang, and P. Buchanan, "Lung cancer: Biology and treatment options," *Biochimica Et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1856, pp. 189–210, Aug. 2015.

[4] K. Funai et al., "Clinicopathologic characteristics of peripheral squamous cell carcinoma of the lung," *The American Journal of Surgical Pathology*, vol. 27, pp. 978–984, Jun. 2003.

[5] N. Zhang et al., "The egfr pathway is involved in the regulation of pd-l1 expression via the il-6/jak/stat3 signaling pathway in egfr-mutated non-small cell lung cancer," *International Journal of Oncology*, vol. 49, pp. 1360–1368, Jul. 2016.

[6] N. Karachaliou, M. F. Bruno, J. W. P. Bracht, and R. Rosell, "Egfr first- and second-generation tkis—there is still place for them in egfr-mutant nsclc patients," *Translational Cancer Research*, vol. 8, p. S23–S47, Jan. 2018.

[7] L. Hajri et al., "Anticancer activity of encapsulated pearl millet polyphenol-rich extract against proliferating and non-proliferating breast cancer cells in vitro," *Cancers*, vol. 16, p. 1750, Apr. 2024.

[8] R. A. Tahir et al., "Pharmacoinformatics and molecular docking reveal potential drug candidates against schizophrenia to target taar6," *Journal of Cellular Physiology*, vol. 234, p. 13263–13276, Dec. 2018.

[9] S. A. Sehgal, S. Mannan, and S. Ali, "Pharmacoinformatic and molecular docking studies reveal potential novel antidepressants against neurodegenerative disorders by targeting hspb8," *Drug Design Development and Therapy*, p. 1605, May 2016.

[10] S. M. Abbasi, "Computational rug design and exploration of potent phytochemicals against cancer through in silico approaches," *ResearchGate*, Oct. 2019. [Online]. Available: https://www.researchgate.net/publication/340503550_Computational_drug_design_and_exploration_of_potent_phytochemicals_against_cancer_through_in_silico_approaches.

[11] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, p. 646–674, Mar. 2011.

[12] S. Choudhari, P. C. Mandave, M. Deshpande, P. Ranjekar, and O. Prakash, "Phytochemicals in cancer treatment: From preclinical studies to clinical practice," *Frontiers in Pharmacology*, vol. 10, Jan. 2020.

[13] M. J. Thun, J. O. DeLancey, M. M. Center, A. Jemal, and E. M. Ward, "The global burden of cancer: priorities for prevention," *Carcinogenesis*, vol. 31, p. 100–110, Nov. 2009.

[14] Madhavan, P. Sankar, K. Ravindranath, R. Soundhararajan, and H. Srinivasan, "Screening the efficacy of compounds from ghee to control cancer: An in silico approach," *Biointerface Research in Applied Chemistry*, vol. 11, p. 14115–14126, Mar. 2021.

[15] E. J. Devlin, L. A. Denson, and H. S. Whitford, "Cancer treatment side effects: A meta-analysis of the relationship between response expectancies and experience," *Journal of Pain and Symptom Management*, vol. 54, pp. 245–258.e2, Aug. 2017.

[16] M. M. J. Moideen et al., "Application of the box–behnken design for the production of soluble curcumin: Skimmed milk powder inclusion complex for improving the treatment of colorectal cancer," *Food Science Nutrition*, vol. 8, p. 6643–6659, Oct. 2020.

[17] M. G. Mokwena, C. A. Kruger, M.-T. Ivan, and A. Heidi, "A review of nanoparticle photosensitizer drug delivery uptake systems for photodynamic treatment of lung cancer," *Photodiagnosis and Photodynamic Therapy*, vol. 22, p. 147–154, Mar. 2018.

[18] D. M. Ornitz and N. Itoh, "The fibroblast growth factor signaling pathway," *Wiley Interdisciplinary Reviews. Developmental Biology*, vol. 4, p. 215–266, Mar. 2015.

[19] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, p. 7–30, Jan. 2020.

[20] H. Sung et al., "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, p. 209–249, Feb. 2021.

[21] N. Duma, R. Santana-Davila, and J. R. Molina, "Non–small cell lung cancer: Epidemiology, screening, diagnosis, and treatment," *Mayo Clinic Proceedings*, vol. 94, pp. 1623–1640, Aug. 2019.

[22] S. H. Bradley, M. P. T. Kennedy, and R. D. Neal, "Recognising lung cancer in primary care," *Advances in Therapy*, vol. 36, pp. 19–30, Nov. 2018.

[23] W. D. Travis et al., "The 2015 world health organization classification of lung tumors," *Journal of Thoracic Oncology*, vol. 10, pp. 1243–1260, Aug. 2015.

[24] W. D. Travis, E. Brambilla, A. P. Burke, A. Marx, and A. G. Nicholson, "Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart," *Journal of Thoracic Oncology*, vol. 10, pp. 1240–1242, Aug. 2015.

[25] K. Takamochi et al., "Novel biomarkers that assist in accurate discrimination of squamous cell carcinoma from adenocarcinoma of the lung," *BMC Cancer*, vol. 16, Sep. 2016.

[26] L. M. Sholl et al., "Liquid biopsy in lung cancer: A perspective from members of the pulmonary pathology society," *Archives of Pathology Laboratory Medicine*, vol. 140, pp. 825–829, May 2016.

[27] H.-J. Oh et al., "Clinical applications of circulating biomarkers in non-small cell lung cancer," *Frontiers in Cell and Developmental Biology*, vol. 12, Aug. 2024.

[28] H. Lemjabbar-Alaoui, O. U. Hassan, Y.-W. Yang, and P. Buchanan, "Lung cancer: Biology and treatment options," *Biochimica Et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1856, pp. 189–210, Aug. 2015.

[29] S. Lutfiya, S. Priya, M. A. P. Manzoor, and S. Hemalatha, "Molecular docking and interactions between vascular endothelial growth factor (VEGF) receptors and phytochemicals: An in-silico study," *Biocatalysis and Agricultural Biotechnology*, vol. 22, p. 101424, Nov. 2019.

[30] Y. Liu et al., "Cancer progression is mediated by proline catabolism in non-small cell lung cancer," *Oncogene*, vol. 39, pp. 2358–2376, Jan. 2020.

[31] M. Atanasova and A. Whitty, "Understanding cytokine and growth factor receptor activation mechanisms," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 47, pp. 502–530, Oct. 2012.

[32] F. Cappuzzo et al., "Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non–small-cell lung cancer," *Journal of the National Cancer Institute*, vol. 97, pp. 643–655, May 2005.

[33] W. Pao and J. Chmielecki, "Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer," *Nature Reviews. Cancer*, vol. 10, pp. 760–774, Oct. 2010.

[34] R. Chen et al., "Emerging therapeutic agents for advanced non-small cell

lung cancer," *Journal of Hematology Oncology*, vol. 13, May 2020.

[35] E. Chowdhury and S. Tiash, "Growth factor receptors: promising drug targets in cancer," *Journal of Cancer Metastasis and Treatment*, vol. 1, p. 190, Jan. 2015.

[36] A. Bennasroune, D. Gardin, G. Aunis, P. Crémel, and P. Hubert, "Tyrosine kinase receptors as attractive targets of cancer therapy," *Critical Reviews in Oncology/Hematology*, vol. 50, pp. 23–38, Apr 2004.

[37] M. Johnson, M. C. Garassino, T. Mok, and T. Mitsudomi, "Treatment strategies and outcomes for patients with EGFR-mutant non-small cell lung cancer resistant to EGFR tyrosine kinase inhibitors: Focus on novel therapies," *Lung Cancer*, vol. 170, pp. 41–51, May 2022.

[38] B. Yin, D.-M. Fang, X.-L. Zhou, and F. Gao, "Natural products as important tyrosine kinase inhibitors," *European Journal of Medicinal Chemistry*, vol. 182, p. 111664, Aug 2019.

[39] T. Kosaka, Y. Yatabe, H. Endoh, H. Kuwano, T. Takahashi, and T. Mitsudomi, "Mutations of the epidermal growth factor receptor gene in lung cancer," *Cancer Research*, vol. 64, pp. 8919–8923, Dec 2004.

[40] P. Bunnjr and W. Franklin, "Epidermal growth factor receptor expression, signal pathway, and inhibitors in non–small cell lung cancer," *Seminars in Oncology*, vol. 29, pp. 38–44, Jan 2002.

[41] N. Hanna et al., "Systemic therapy for stage IV non–small-cell lung cancer: American society of clinical oncology clinical practice guideline update," *Journal of Clinical Oncology*, vol. 35, pp. 3484–3515, Oct 2017.

[42] S. Novello et al., "Metastatic non-small-cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 27, pp. v1–v27, Sep 2016.

[43] D. A. E. Cross et al., "AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer," *Cancer Dis-*

covery, vol. 4, pp. 1046–1061, Sep 2014.

[44] T. Mok et al., "CNS response to osimertinib in patients (pts) with T790M-positive advanced NSCLC: Data from a randomized phase III trial (AURA3)," *Journal of Clinical Oncology*, vol. 35, p. 9005, May 2017.

[45] S. N. Gimba, A. Nanda, and M. A. Karage, "Medicinal chemistry comparative phytochemical screening on three growth stages of Pennisetum pedicellatum Trin," *International Journal of Scientific and Research Publications*, vol. 9, p. p8718, Mar 2019.

[46] K. Jukanti, C. L. L. Gowda, K. N. Rai, V. K. Manga, and R. K. Bhatt, "Crops that feed the world 11. Pearl millet (Pennisetum glaucum L.): An important source of food security, nutrition, and health in the arid and semi-arid tropics," *Food Security*, vol. 8, pp. 307–329, Mar 2016.

[47] S. Deepak, S. Niranjan-Raj, A. Mithöfer, and S. H. Shetty, "Nutritional biofortification in pearl millet," *The European Journal of Plant Science and Biotechnology*, vol. 6, pp. 87–90, Jan 2012.

[48] S. A. Waziri, A. Bala, and H. Shehu, "A comparative analysis of nutrients and mineral elements content of Andropogon gayanus Kunth and Pennisetum pedicellatum Trin," *Nigerian Journal of Basic and Applied Sciences*, vol. 21, Jun 2013.

[49] M. N. Hellen and C. N. Mutuku, "Comparative study on the antibacterial and chemical constituents of Pennisetum glaucum (pearl millet) and Zea mays (maize)," *International Journal of Nutrition and Metabolism*, vol. 7, pp. 46–51, Apr 2015.

[50] A. C. Slama, F. Sakouhi, S. Boukhchina, and L. Radhouane, "Fatty acids, phytochemical composition, and antioxidant potential of pearl millet oil," *Journal of Consumer Protection and Food Safety*, vol. 15, pp. 145–151, Aug 2019.

[51] D. N'Dri, T. Mazzeo, M. Zaupa, R. Ferracane, V. Fogliano, and N. Pellegrini, "Effect of cooking on the total antioxidant capacity and phenolic profile of

some whole-meal African cereals," *Journal of the Science of Food and Agriculture*, vol. 93, pp. 29–36, Aug 2012.

[52] V. S. Nambiar, N. Sareen, M. Daniel, and E. B. Gallego, "Flavonoids and phenolic acids from pearl millet (Pennisetum glaucum) based foods and their functional implications," *Functional Foods in Health and Disease*, vol. 2, p. 251, Jul 2012.

[53] Nani et al., "Effects of polyphenols and lipids from Pennisetum glaucum grains on T-cell activation: Modulation of Ca2+ and ERK1/ERK2 signaling," *BMC Complementary and Alternative Medicine*, vol. 15, Dec 2015.

[54] O. A. Ojo et al., "Phytochemical properties and pharmacological activities of the genus Pennisetum: A review," *Scientific African*, vol. 16, p. e01132, Feb 2022.

[55] Y. Liang, T. Zhang, and J. Zhang, "Natural tyrosine kinase inhibitors acting on the epidermal growth factor receptor: Their relevance for cancer therapy," *Pharmacological Research*, vol. 161, p. 105164, Aug 2020.

[56] R. Muthuvel, S. Jagannathan, N. K. Pariyapurath, R. G. Pachamuthu, M. Mathanmohun, and S. Sagadevan, "Harnessing nutritional powerhouse: Millets and probiotics in anticancer therapy," *Current Pharmacology Reports*, Jul 2024.

[57] K. K. Chaudhary and N. Mishra, "A review on molecular docking: Novel tool for drug discovery," 2016. Available online: https://www.jscimedcentral.com/jounal-article-info/JSM-Chemistry/A-Review-on-Molecular-Docking%3A-Novel-Tool-for-Drug-Discovery-8883#.

[58] Y. Liu, M. Grimm, W.-T. Dai, M.-C. Hou, Z.-X. Xiao, and Y. Cao, "CB-Dock: A web server for cavity detection-guided protein–ligand blind docking," *Acta Pharmacologica Sinica*, vol. 41, no. 1, pp. 138–144, 2019.

[59] R. Santhoshkumar and A. Yusuf, "In silico structural modeling and analysis of physicochemical properties of curcumin synthase (Curs1, Curs2, and Curs3)

proteins of *Curcuma longa*," *Journal of Genetic Engineering and Biotechnology*, vol. 18, no. 1, 2020.

[60] M. Blum et al., "The InterPro protein families and domains database: 20 years on," *Nucleic Acids Research*, vol. 49, no. D1, pp. D344–D354, 2020.

[61] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "PubChem: A public information system for analyzing bioactivities of small molecules," *Nucleic Acids Research*, vol. 37, no. Web Server, pp. W623–W633, 2009.

[62] S. Kim and E. E. Bolton, "PubChem: A large-scale public chemical database for drug discovery," in *Methods and Principles in Medicinal Chemistry*, pp. 39–66, 2023.

[63] K. Roy, S. Kar, and R. N. Das, "Computational chemistry," in *Elsevier eBooks*, pp. 151–189, 2015.

[64] T. I. Adelusi et al., "Molecular modeling in drug discovery," *Informatics in Medicine Unlocked*, vol. 29, 2022.

[65] M. Agostino, R. L. Mancera, P. A. Ramsland, and E. Yuriev, "AutoMap: A tool for analyzing protein–ligand recognition using multiple ligand binding modes," *Journal of Molecular Graphics and Modelling*, vol. 40, pp. 80–90, 2013.

[66] M. Ifran, H. Mudalige, and O. Perera, "Protein-ligand docking study for the identification of binding sites and ligands against the ischemic stroke receptors," 2022.

[67] X. Chen, H. Li, L. Tian, Q. Li, J. Luo, and Y. Zhang, "Analysis of the physicochemical properties of acaricides based on Lipinski's rule of five," *Journal of Computational Biology*, vol. 27, no. 9, pp. 1397–1406, 2020.

[68] J. G. M. Mvondo, A. Matondo, D. T. Mawete, S.-M. N. Bambi, B. M. Mbala, and P. O. Lohohola, "In silico ADME/T properties of quinine derivatives using SwissADME and PKCSM webservers," *International Journal of Tropical*

Disease & Health, pp. 1–12, 2021.

[69] Biharee et al., "Flavonoids as promising anticancer agents: An in silico investigation of ADMET, binding affinity by molecular docking and molecular dynamics simulations," *Journal of Biomolecular Structure and Dynamics*, vol. 41, no. 16, pp. 7835–7846, 2022.

[70] M. Kanehisa et al., "KEGG for linking genomes to life and the environment," *Nucleic Acids Research*, vol. 36, no. Database, pp. D480–D484, 2007.

[71] F. Stanzione, I. Giangreco, and J. C. Cole, "Use of molecular docking computational tools in drug discovery," *Progress in Medicinal Chemistry*, pp. 273–343, 2021.

[72] H. Prabhavathi, K. R. Dasegowda, K. H. Renukananda, P. Karunakar, K. Lingaraju, and H. R. Naika, "Molecular docking and dynamic simulation to identify potential phytocompound inhibitors for EGFR and HER2 as anti-breast cancer agents," *Journal of Biomolecular Structure and Dynamics*, vol. 40, no. 10, pp. 4713–4724, 2020.

[73] M. I. Hosen et al., "In-silico approach to characterize the structure and function of a hypothetical protein of monkeypox virus exploring Chordopox-A20R domain-containing protein activity," *Antiviral Therapy*, vol. 29, no. 3, 2024.

[74] S. C. Moldoveanu and V. David, "Properties of analytes and matrices determining HPLC selection," in *Elsevier eBooks*, pp. 189–230, 2017.

[75] D. G. Gamage, A. Gunaratne, G. R. Periyannan, and T. G. Russell, "Applicability of instability index for in vitro protein stability prediction," *Protein and Peptide Letters*, vol. 26, no. 5, pp. 339–347, 2019.

[76] S. Panda and G. Chandra, "Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates," *Bioinformation*, vol. 8, no. 18, pp. 891–896, 2012.

[77] H. Duan, A. Er-Bu, Z. Dongzhi, H. Xie, B. Ye, and J. He, "Alkaloids from

Dendrobium and their biosynthetic pathway, biological activity and total synthesis," *Phytomedicine*, vol. 102, 2022.

[78] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.

[79] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, and T. Gray, "How to measure and predict the molar absorption coefficient of a protein," *Protein Science*, vol. 4, no. 11, pp. 2411–2423, 1995.

[80] R. D. Requião, L. Fernandes, H. J. A. D. Souza, S. Rossetto, T. Domitrovic, and F. L. Palhano, "Protein charge distribution in proteomes and its impact on translation," *PLoS Computational Biology*, vol. 13, no. 5, 2017.

[81] D. A. Korasick and J. M. Jez, "Protein domains: structure, function, and methods," in *Elsevier eBooks*, pp. 91–97, 2015.

[82] S. Hunter et al., "InterPro: the integrative protein signature database," *Nucleic Acids Research*, vol. 37, pp. D211–D215, Oct. 2008.

[83] E. H. Bowler-Barnett, J. Fan, J. Luo, M. Magrane, M. J. Martin, and S. Orchard, "UniProt and mass spectrometry-based proteomics—a 2-way working relationship," *Molecular Cellular Proteomics*, vol. 22, p. 100591, Jun. 2023.

[84] C. Wallace, R. A. Laskowski, and J. M. Thornton, "LigPlot: a program to generate schematic diagrams of protein-ligand interactions," *Protein Engineering Design and Selection*, vol. 8, pp. 127–134, Jan. 1995.

[85] R. Roskoski, "The ERBB/HER family of protein-tyrosine kinases and cancer," *Pharmacological Research*, vol. 79, pp. 34–74, Nov. 2013.

[86] H. Duan, A. Er-Bu, Z. Dongzhi, H. Xie, B. Ye, and J. He, "Alkaloids from Dendrobium and their biosynthetic pathway, biological activity and total synthesis," *Phytomedicine*, vol. 102, p. 154132, Apr. 2022.

[87] L. Guan et al., "ADMET-score – a comprehensive scoring function for evaluation of chemical drug-likeness," *Medicinal Chemistry Communications*, vol.

10, pp. 148–157, Nov. 2018.

[88] M. M. Alsanosi, C. Skiffington, and S. Padmanabhan, "Pharmacokinetic pharmacogenomics," in *Elsevier eBooks*, pp. 341–364, 2014.

[89] K. A. Azzam, "SwissADME and pkCSM webservers predictors: an integrated online platform for accurate and comprehensive predictions for in silico ADME/T properties of artemisinin and its derivatives," *Kompleksnoe Ispolzovanie Mineralnogo Syra = Complex Use of Mineral Resources*, vol. 325, pp. 14–21, Nov. 2022.

[90] Y. Yeni and R. A. Rachmania, "The prediction of pharmacokinetic properties of compounds in *Hemigraphis alternata* (Burm.f.) T. Ander leaves using pkCSM," *Indonesian Journal of Chemistry*, vol. 22, p. 1081, Jul. 2022.

[91] D. A. M. Adnyaswari, A. W. Indrayani, and I. G. A. Artini, "In silico toxicity and pharmaceutical properties to get candidates for antitumor drug," *Journal of Pharmaceutical Research International*, vol. 36, pp. 1–11, Feb. 2024.

[92] O. K. Didigwu and C. O. Nnadi, "Theoretical drug-likeness, pharmacokinetic and toxicities of phytotoxic terpenoids from the toxic plants-phytotoxins," *Tropical Journal of Natural Product Research*, vol. 8, Nov. 2024.

[93] P. D. Utami, V. Yudo, and R. Budiarti, "Blockade of multiple pathways of *P. falciparum* by quinoxaline from curry fish (*S. hermanni*) using an in silico approach," *Indian Journal of Pharmaceutical Education and Research*, vol. 59, pp. 326–337, Jan. 2025.

[94] A. M. Rahman, H. M. Korashy, and M. G. Kassem, "Gefitinib," in *Profiles of Drug Substances, Excipients, and Related Methodology*, pp. 239–264, 2014.

[95] R. B. Semwal, D. K. Semwal, S. Combrinck, J. Trill, S. Gibbons, and A. Viljoen, "Acacetin—a simple flavone exhibiting diverse pharmacological activities," *Phytochemistry Letters*, vol. 32, pp. 56–65, May 2019.